FEDERAL RESERVE BANK *of* ATLANTA WORKING PAPER SERIES

The Effect of Large Investors on Asset Quality: Evidence from Subprime Mortgage Securities

Manuel Adelino, W. Scott Frame, and Kristopher S. Gerardi

Working Paper 2014-4a
Revised March 2017

Abstract: The government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac—the dominant investors in subprime mortgage-backed securities before the 2008 crisis—substantively affected collateral composition in this market. Mortgages included in securities designed for the GSEs performed better than those backing other securities in the same deals, holding observable risk constant. Consistent with the transmission of private information, these effects are concentrated in low-documentation loans and for issuers that were highly dependent on the GSEs and were corporate affiliates of the mortgage originators. Additional analysis of yield spreads shows that these performance differences were not reflected in prices.

JEL classification: G17, G21, G23

Key words: mortgage default, GSEs, securitization, private information

---

For helpful comments and discussions, the authors thank Sumit Agarwal, Brent Ambrose, Edward Coulson, Andra Ghent, Robert Hansen, Peter Niculescu, Kiat Ying Seah, Shane Sherlund, Larry Wall, Paul Willen, and Vincent Yao as well as seminar participants at AEA, AREUEA, Bangor University, University of Cambridge, Cornell University, Georgia State University, Kellogg Junior Finance Conference, New York University, NUS-IRES Symposium, Southern Finance Association, Syracuse University, University of Juan Carlos III, and the University of Kansas. They also thank Neil Desai for outstanding research assistance. Frame also thanks the University of North Carolina at Charlotte and the Federal Reserve Bank of Richmond for financial support. This paper was previously circulated under the title “The Role of Blockholders in Debt Markets: Fannie Mae, Freddie Mac and Subprime MBS.” The views expressed here are the authors’ and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the authors’ responsibility.

Please address questions regarding content to Manuel Adelino, Duke University, Fuqua School of Business, 100 Fuqua Drive, Durham, NC 27708, 919-660-7981, manuel.adelino@duke.edu; Scott Frame, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8783, scott.frame@atl.frb.org; or Kristopher Gerardi, Federal Reserve Bank of Atlanta, Research Department, 1000 Peachtree Street NE, Atlanta, GA 30309-4470, 404-498-8561, kristopher.gerardi@atl.frb.org.

1 Introduction

Informational frictions and conflicts of interest between issuers of financial claims and the owners of those claims are central topics in financial economics. It is well understood that the presence of large, concentrated claimholders can have significant effects on managerial incentives, which, in turn, can positively affect firm value and market outcomes.[1] However, Shleifer and Vishny (1997) note that large blockholders can also impose costs if they are able to promote their own interests at the expense of other shareholders. On the debt-side, banks incur screening and monitoring costs to acquire private information about borrowers, which serves to reduce adverse selection and moral hazard problems (Boot (2000)). But relationship banks may exploit this informational advantage to "lock-in" borrowers and extract rents from them over time (e.g., Sharpe (1990), Degryse and Van Cayseele (2000), and Schenone (2010)).

The literature on the influence of concentrated investors has focused almost exclusively on equity markets and relationship-based bank lending markets: both represent risky, informationally-sensitive financial claims. Investors in informationally insensitive arms-length debt markets, on the other hand, are generally thought of as more passive providers of capital. This paper studies the outsized role played by the government-sponsored enterprises (GSEs), Fannie Mae and Freddie Mac, in the subprime private-label mortgage-backed securities (PLS) market during the recent housing boom.[2] It analyzes security-level and underlying loan-level data and documents that the GSEs received preferential treatment from PLS security issuers along two dimensions: loan quality and security pricing.

The participation of the GSEs in the PLS market provides a unique setting for studying whether investor concentration matters in arms-length debt markets for several reasons.

[1]See Shleifer and Vishny (1986), Brav et al. (2008), Burkart et al. (1997), Edmans (2009), McCahery et al. (2016) and Parrino et al. (2003).

[2]In addition to investing in PLS, Fannie Mae and Freddie Mac play a central role in the U.S. mortgage market through their "credit guarantee" or "securitization" businesses, whereby mortgage originators exchange portfolios of mortgages for securities that represent an interest in the same pool. In this case, the GSEs agree to ensure timely payment of principal and interest on the securities in exchange for a monthly insurance premium ("guarantee fee").

First, Fannie Mae and Freddie Mac together purchased roughly 30 percent of the total dollar volume of subprime PLS issued during the recent U.S. housing boom (2003–2007), making them by far the largest single investors in this asset class.[3] Second, securities purchased by the GSEs can be identified by exploiting the fact that, by law, Fannie Mae and Freddie Mac are only allowed to acquire mortgages with principal balances below the conforming loan limit, a fixed dollar amount set annually by the government. This prompted the creation of subprime PLS deals designed specifically for GSE participation. Those deals included securities that were backed only by mortgages below the conforming loan limit (CLL), as well as other securities (issued at the same time, by the same issuer, and with the same credit enhancement and credit rating) that were backed by both mortgages below the CLL and jumbo mortgages (those with a balance above the CLL) and could not be purchased by the GSEs. It has, in general, proven difficult to establish who the final holders of individual PLS were, and this paper is one of the first to do this for a large segment of the market.[4] (Indeed, it is very difficult to match investors with individual security issues in any fixed income market.) Third, this institutional arrangement permits the analysis of selection problems related to the endogenous matching of large claimholders with securities issuers. The identification strategy used in this paper compares the performance of mortgages backing securities bought by the GSEs with those issued as part of the same deal purchased by other investors. Finally, there are detailed collateral characteristics available in the mortgage-backed securities market that are generally not present in other debt markets. This in turn permits the identification of the role of issuer private information in the allocation of risk between securities sold to the GSEs and those sold to other investors.

Using within-deal variation, and controlling for all relevant observable underwriting characteristics, this study finds that mortgages in GSE pools (i.e., those backing securities pur-

[3]Given their focus on the highest rated tranches, the GSE share of triple-A subprime PLS actually approached 40 percent during this period. Ghent et al. (2015) analyze a random sample of 100 PLS prospectus supplements and find a similar market share for the GSEs.

[4]One exception is the use of insurance company trading data, as in Merrill et al. (2014) and Ellul et al. (2014).

chased by the GSEs) performed significantly better than mortgages backing other securities. The within-deal setup rules out alternative explanations for the results, including effects due to macro conditions at the time of issuance and unobserved heterogeneity at the level of the issuer and originator, since these are all shared by loans in the same deal. In addition, the split between GSE and non-GSE loan pools applied exclusively to the triple-A securities in these deals, while lower-rated tranches received cash flows from both pool types and provided credit support for all senior securities (GSE and non-GSE triple-A alike). This removes any concern that differences in subordination, credit support, or risk retention could explain the results.

The results are most consistent with a private information mechanism by which concentrated investors use their market power to obtain preferential treatment, likely at the expense of other, less influential investors. Consistent with the issuers using private information, this study shows that the difference in performance between mortgages in GSE and non-GSE pools comes primarily from low-documentation loans, where soft information has been found to be especially important (Keys et al. (2012); Saengchote (2013); Jiang et al. (2014a); Jiang et al. (2014b); Begley and Purnanandam (2017)). The difference in default rates between GSE and non-GSE pools for low-documentation mortgages is 3.9 percentage points, whereas the difference for full-documentation loans is only 0.7 percentage points. This is consistent with the idea that issuer sorting of mortgages across pools within a deal was the source of the higher-quality mortgages in the GSE pools.

To identify a private information mechanism, this study constructs a measure of the frequency of interactions between issuers and the GSEs over time. Information on the identity of subprime PLS issuers is used to compute the fraction of previous deals for each issuer that included GSE pools. Issuers that often include securities designed for Fannie Mae and Freddie Mac in their deals are likely to wish to maintain good standing with the two GSEs for future business. Deals arranged by issuers that frequently include GSE pools in their offerings exhibit significantly larger differences between GSE and non-GSE pool performance.

Furthermore, the (within-deal) performance differences are fully explained by differences in low-documentation loans and by the fraction of previous deals with GSE participation.

We further build on these tests by using the identity of the issuers and separating deals into those where the issuer of the securities and the originator of the underlying mortgages were affiliated (i.e., either the same institution or part of the same vertically integrated entity) and those where they were not. Following previous literature (Demiroglu and James (2012) and Furfine (2014)), the rationale for the test is that issuers that are affiliated with the mortgage originators are more likely to have private information about unobserved characteristics of groups of borrowers. The transmission of this private information could take various forms. First, mortgage originators could hold soft information about individual loan quality that they are able to transmit directly to issuers. Second, and perhaps more likely, mortgage lenders may hold private information about the origination practices of certain brokers, or information about types of borrowers (e.g., the quality of income reporting) for groups of loans that were being securitized, that can more easily transmit to affiliated issuers. Consistent with the previous tests, the results indicate that the superior performance of low-documentation loans in GSE pools and of loans securitized by issuers with frequent interactions with the GSEs is especially strong in deals where the issuer and the originator of the mortgages were affiliated institutions.

The baseline (within-deal) models are highly saturated with a large set of categorical controls, but there might still be a concern that covariate imbalance could be responsible for some of the results. To account for this, the empirical specifications are reestimated on a matched sample of loans. Using a nearest-neighbor match based on estimated propensity scores, we match loans in GSE and non-GSE pools within the same deal. This procedure significantly reduces observable differences across treatment and control loans, and the results show that the performance differences are largely unchanged in both statistical significance and economic magnitude. In addition, a battery of robustness checks are conducted to ensure that omitted variable bias is not driving the results.

We complement the loan performance results with an analysis of PLS yield spreads at-issuance for those triple-A securities that were either floating-rate or inverse floaters. Although information on the actual prices paid for the securities are not available, restricting the analysis to floating-rate securities minimizes the likelihood that they were not issued at par. Yield spreads associated with GSE floating-rate PLS are, on average, 3-6 basis points higher than those associated with non-GSE securities after controlling for weighted-average pool characteristics and deal fixed effects. This difference is consistent with the results from the ex-ante credit risk analysis, in which GSE loan pools were found to be riskier than non-GSE pools based on information at the time of issuance. Furthermore, the results show that the higher GSE yield spreads are driven by issuers who have been previously dependent on the GSEs for business, which is suggestive of a GSE catering effect. Altogether, this analysis suggests that prices reflected higher ex-ante credit risk associated with GSE loan pools but not their superior ex-post performance, providing further support for a private information interpretation.

As a final exercise, this study exploits between-deal variation in loan performance and pricing to address the question of whether investors in deals with a presence by the GSEs were harmed relative to investors in deals where the GSEs did not participate. Specifically, it compares the performance of non-GSE loan pools in deals with and without GSE participation as well as their corresponding triple-A yield spreads. The results show that non-GSE loans in deals without GSE participation were between 2 and 3 percentage points less likely to default than similar loans in deals with GSE participation, while their corresponding triple-A yield spreads were slightly higher (approximately 1–2 basis points). This finding is consistent with the GSEs' presence having a negative effect on other PLS investors in the market.

The message from this paper is that the presence of concentrated claimholders influences security performance and pricing, even in the case of informationally insensitive (highly rated) arm's-length debt markets. The results suggest a GSE catering effect, potentially owing to reputational concerns on the part of PLS issuers or due to the GSEs' perceived higher likelihood of enforcing contractual clauses (as they are more likely to internalize both the benefits and costs of ex-post monitoring).

2 PLS deal structure and identification strategy

A key input to this paper is the categorization of PLS securities into those that were specifically structured to allow Fannie Mae and Freddie Mac to invest in them and those that were not. To do this, we use the fact that the GSEs were only allowed to purchase triple-A rated securities backed by mortgages below the conforming loan limit (CLL).[5] In the PLS market this restriction was overcome by using more than one collateral pool for the same deal: one pool would include only mortgages with a balance below the CLL and one or more other pools included a mix of mortgages below the CLL, and "jumbo" loans (those above the limit). Figure 1 provides a stylized illustration of how a typical subprime PLS deal with GSE participation was structured. PLS issuers created triple-A securities intended for the GSEs that received cash flows only from the conforming loan pool, and other triple-A securities that received cash flows from the other pool(s) for other investors. Remaining cash flows from all pools in a deal accrued to the subordinate/junior tranches (double-A rated and below). Other than the requirement that all loans be below the CLL, no other underwriting restrictions differentiated loans that collateralized GSE PLS securities from those that collateralized PLS securities purchased by other investors.

Considering this institutional feature of the subprime PLS market, a simple algorithm is designed to identify mortgage pools that backed securities eligible to be purchased by either Fannie Mae or Freddie Mac between 2003 and 2007. A pool is classified as being a "GSE pool" if at least 99 percent of loans in the pool are below the CLL at the time of issuance to investors. Crucially for this identification strategy, the only reason why an

[5]The CLL for most single-family homes between 2006 and 2013 was $417,000. Single-family homes in certain high-cost areas have higher limits, as do two-to-four family properties. We refer the reader to the Online Appendix for a list of single-family conforming loan limits during our sample period.

issuer would structure a deal in this way would be to attract the GSEs as investors, since no other investors in the market are constrained by the CLL. To illustrate the validity of this approach, we identify 476 of the 478 subprime PLS securities included in lawsuits brought by the Federal Housing Finance Agency (FHFA) in 2011 alleging fraudulent marketing and sales materials for GSE-purchased PLS.[6] Additional details about the algorithm and this validation exercise are found in the Online Appendix.

Results from this simple algorithm shows that the GSEs invested heavily in triple-A subprime PLS from 2003 through 2007, accounting for approximately one-third of overall issuance during this period. There are three potential reasons why Fannie Mae and Freddie Mac were interested in investing so heavily in the subprime PLS market. First, the GSEs were limited, by law, to operating exclusively in the secondary conforming mortgage market. Second, the two firms benefitted from an implicit federal guarantee of their debt obligations that resulted in a significant funding advantage and little market discipline (e.g., Greenspan (2005)). Hence, taken together, Fannie Mae and Freddie Mac faced strong incentives to grow and acquire all eligible mortgage-related assets with yields above their cost of funding. Third, mortgages funded through PLS deals could be counted against the GSEs' affordable housing goals (in proportion to their investment).[7] Analysis by the U.S. Department of Housing and Urban Development illustrates the goal-richness of subprime PLS acquired by Fannie Mae and Freddie Mac in 2004 and 2005 (Bunce (2007)).

The buyers of triple-A PLS securities with claims on the "non-GSE" pools were much more dispersed. According to Greenlaw et al. (2008), there were seven main groups of non-GSE investors in subprime PLS before 2007: commercial banks, investment banks, insurance companies, hedge funds, finance companies, mutual funds, and pension funds. For various

[6]Details on the lawsuits are available on the FHFA website at http://www.fhfa.gov/Media/PublicAffairs/Pages/FHFAs-Update-on-Private-Label-Securities-Actions.aspx

[7]As Robert Levin, the former chief business officer of Fannie Mae, told the Federal Crisis Inquiry Commission, buying private-label mortgage-backed securities "was a money making activity—it was all positive economics. . . . [T]here was no trade-off [between making money and hitting goals], it was a very broad-brushed effort" that could be characterized as "win-win-win: money, goals, and share" (FCIC (2011)). Ghent et al. (2015), on the other hand, argue that goals are unlikely to have been a primary motivation for these purchases.

regulatory reasons, U.S. and foreign commercial banks, investment banks, and insurance companies were the most likely investors (on and off balance sheet) in non-GSE triple-A subprime securities. For the purposes of this paper, the key assumption is that these investors were much more dispersed than the GSEs.

2.1 Empirical Specification

The study's primary empirical analysis involves comparing the default rates of mortgages in GSE pools with those in non-GSE pools to determine whether there are systematic differences in loan quality that were unobservable to PLS investors at the time of contracting. This is a unique setting that permits the isolation of the impact of a specific investor (or class of investors) on the quality of the underlying assets and on the behavior of issuers.

Performance differences due to observable and unobservable factors are separated by conditioning on a large set of loan and borrower characteristics used in most mortgage performance models. These characteristics were readily available to institutional investors. After documenting differences in conditional default rates between GSE and non-GSE loan pools, a battery of tests is conducted to determine whether the performance differences are driven by private information that is unobservable to other PLS investors in the market.

The study's main identification strategy relies on comparing the ex-post performance of GSE and non-GSE mortgage pools *in the same deal*, after controlling for the observable risk characteristics of the loans. This allows all deal-level characteristics, such as issuer and originator, to be excluded as drivers of the results, since these entities were all shared by loans in the same deal. This identification strategy also controls for potential differences in subordination levels across GSE and non-GSE triple-A securities.[8]

Specifically, the paper estimates loan-level regressions of the form

$$LHS_{ijz} = \alpha + \beta_1 X_{ijz} + \beta_2 GSE_{iz} + \eta_j + \varepsilon_{ijz}, \tag{1}$$

[8] He et al. (2012) report significant variation in subordination levels across PLS deals issued between 2000 and 2006 (see Table 1 in their paper).

where z identifies each mortgage in pool i within deal j. LHS_{ijz} is a measure of mortgage delinquency (described in the next section). X_{ijz} is a vector of mortgage-level control variables that includes all relevant observable borrower, loan, and geographic characteristics, and GSE_{iz} is an indicator variable that is equal to 1 for mortgages in GSE pools and 0 otherwise. The term η_j represents deal-level fixed effects. Note that the inclusion of deal-level fixed effects accounts for very fine time fixed effects (effectively one fixed effect for each date that an issuance is observed). Standard errors are clustered by quarter of issuance.

The vector of mortgage-level control variables includes the combined loan-to-value (CLTV) ratio, the natural logarithm of the original loan balance, the original interest rate, the credit score, the original term, the number of months between origination and issuance (seasoning), and indicator variables for first-lien loans, low-documentation loans, interest-only loans, balloon loans, negative amortization loans, residence status (owner-occupied, investor/vacation home), loan purpose (cash-out refinance, other refinance, purchase), property type (condominium, multi-family, single-family), and the existence of a prepayment penalty. Also included are the county-level unemployment rate and the level of the house price index at the time of issuance (normalized by setting the index value for January 2000 to 100 for each county), including the changes in these series from the time of issuance through the end of the default horizon, as well as a full set of state-level fixed effects. Additional indicator variables are included whenever there are missing observations for any of the controls.

The analysis of performance after origination is supplemented with an analysis of ex-ante credit risk. Specifically, we construct and compare ex-ante expected default probabilities between loans in GSE and non-GSE pools, conditional on observable information, to determine if there were significant differences in expected performance that could be forecast by investors in real time. Finally, PLS security yield spreads at issuance are analyzed for evidence of any systematic differences in prices between GSE and non-GSE securities that might reflect differences in ex-ante credit risk known to investors.

3 Data and summary statistics

The data used in this paper come from two main sources. All loan-level data come from CoreLogic's private-label securities database, which covers virtually the entire PLS market. This dataset contains information on the underwriting characteristics of the loans underlying the mortgage-backed securities at origination (as discussed above) and the performance of the loans from the month of origination through 2012. Importantly, this dataset includes an identifier for the mortgage pool that each loan belongs to, which permits pool-level variables to be constructed using individual loan data (as opposed to datasets that just include the deal to which a loan belongs, which does not allow for the distinction between GSE and non-GSE mortgage pools used here). In addition, the dataset contains security identifiers (CUSIPs) and deal identifiers.

The raw CoreLogic data sample includes 13,189,213 mortgages that backed subprime PLS issued between 2000 and 2007. Excluded are loans that are not first or second liens (22,395 loans), loans that were seasoned more than 12 months at the time of issuance (813,901 loans), and loans originated in the U.S. territories of Puerto Rico and the Virgin Islands (126 loans). This leaves 12,352,791 loans in 3,987 mortgage pools that collateralized 2,161 subprime deals issued between 2000 and 2007. The empirical analysis restricts attention primarily to loans backing deals issued between 2003 and 2007 for two reasons. First, the majority of securities issued in 2003 (and later) have not been fully paid off (or even close to paid off) by the third quarter of 2007, when the mortgage default and foreclosure crisis begins. Second, the GSEs did not become major investors in the subprime PLS market until 2003 (see Table 1, discussed in greater detail below).

Additional information on subprime PLS attributes was hand-collected from Bloomberg. The data fields include security identifiers (including CUSIP and ticker), issuer name, issuance date, the identification of the loan pool that the security has claims on, the spread over one-month LIBOR at origination, and the weighted average life as advertised in the prospectus. The dataset we obtain from Bloomberg covers over 90 percent of all subprime

PLS issued in the United States between 2000 and 2007. The CoreLogic and Bloomberg datasets are combined by merging on individual security CUSIPs.

Finally, monthly unemployment rates at the county level are obtained from the Bureau of Labor Statistics, and monthly county-level house price indices are gathered from CoreLogic.

3.1 Summary Statistics

Table 1 displays the aggregate dollar amount of subprime PLS issued each year over the period 2000–2008 (column (1)) obtained from Inside Mortgage Finances 2011 Mortgage Market Statistical Annual. The table shows the rapid growth in the market that took place during the housing boom and the steep decline that occurred at the onset of the financial crisis. In 2000, a little more than $52 billion subprime PLS was issued (74 deals). Amount issued peaked at $465 billion in 2005 (457 deals in total), remained roughly constant in 2006 (with an increase in the number of deals to 523), and then dropped precipitously in 2007 to just over $200 billion. Since 2007 the subprime PLS market has virtually disappeared.

Unfortunately, there is no publicly available information on the exact dollar amount of subprime PLS purchased by Fannie Mae and Freddie Mac over the entire sample period. We were able to obtain this information for 2006–2008 from the FHFA's 2011 Annual Report to Congress, which is displayed in column (3) of the table. Prior to 2006, the report does not break out PLS purchases by type of security (subprime, alt-a, prime) for Freddie Mac. Column (4) of Table 1 shows estimates of the annual amount of subprime PLS purchased by Fannie Mae and Freddie Mac based on the algorithm. To obtain these totals, all triple-A securities classified as being eligible for GSE purchase and included in Bloomberg are summed over. The numbers obtained for 2006–2008 are very close to the FHFA figures, which suggests that the algorithm is truly identifying mortgage pools backing subprime PLS purchased by the GSEs. Column (5) displays the GSEs' combined subprime PLS market shares based on aggregate amounts derived from the algorithm. In 2001, Fannie Mae and Freddie Mac purchased less than 4 percent of total subprime PLS issued, but in 2004 they bought almost

40 percent. Given that the GSEs purchased exclusively triple-A rated securities within the deals, their share of the total triple-A volume of subprime PLS is even higher. It is perhaps even more striking that, starting in 2002, more than half of all subprime PLS deals included securities that were specifically designed for the GSEs (column (6)). This fraction peaked in 2004, when 72 percent of all deals included GSE securities.

Table 2 shows detailed, loan-level summary statistics broken down by mortgages in GSE versus non-GSE pools. Three sets of summary statistics are shown: pooled means for all loans, means after taking into account deal fixed effects, and means for a matched sample of loans. Means conditional on deal fixed effects are computed for GSE and non-GSE pools by running a loan-level regression of each variable of interest (FICO, balance, CLTV ratios, etc.) on a constant, an indicator for GSE pools, and indicator variables for each deal (which absorbs the deal-specific average of each variable). The non-GSE mean shown is the coefficient on the constant for each of these regressions, and the GSE mean is given by the coefficient on the constant plus the coefficient on the GSE variable. This section focuses on the statistics for the full sample (raw and with deal fixed effects); the matched sample is discussed in Section 4.4. The top panel of the table shows information on the means of the continuous variables included in the mortgage performance regressions, while the bottom panel shows means of the dummy variables. There are some notable differences between GSE and non-GSE mortgage pools. For example, average FICO scores, average mortgage terms, and the fractions of purchase mortgages, balloon mortgages, and adjustable-rate mortgages (ARMs) are significantly different across the two types of pools. In general, the differences are significantly ameliorated with the inclusion of deal fixed effects, and, as is discussed below in Section 4.4, are further mitigated when a matched sample of loans is constructed.

The bottom three rows of Table 2 display unconditional average default rates for mortgages in both GSE and non-GSE pools. A mortgage is asssumed to be in default if the borrower is at least two payments behind (60+ days delinquent) at any point between origination and a given point in time. These default rates are calculated directly from the

CoreLogic database using information on loan performance from the time of origination. We use a 60-day delinquency cutoff, rather than a more serious threshold of distress such as the initiation of foreclosure proceedings, so that our default definition reflects borrower behavior that is not confounded by the decisions of servicers. Default rates are reported through the end of 2008, 2010, and 2012 (the end of the Corelogic data sample). The table shows that average default rates associated with GSE pools are significantly lower (by approximately five percentage points for the 2010 and 2012 horizons), and that this difference is still 3–4 percentage points when we compare loans in the same deals.

4 GSE participation and mortgage performance

The first part of the analysis considers the performance after origination of mortgages included in GSE pools relative to those in non-GSE pools. Linear probability models (LPM) are estimated where the dependent variable is equal to one if a loan is in default (i.e., 60 days or more delinquent) between origination and the end of 2008, 2010 or 2012. Two factors related to the inclusion of deal-level fixed effects explain the choice of LPMs rather than non-linear discrete choice models (such as logit or probit). First, the combination of a large sample (over 10 million loans) and roughly 2,000 deal fixed effects makes it computationally difficult to estimate non-linear models. Second, estimating non-linear models with fixed effects can result in inconsistent estimates due to the incidental parameters problem (Neyman and Scott (1948)).[9]

The main set of results for all three horizons is displayed in Table 3. The first column in each horizon panel corresponds to a regression that only includes issue year effects in the covariate set. The estimated GSE coefficient is positive for all three horizons, and the quantitative magnitude of the coefficients implies that loans backing GSE pools have default rates that are on average 120 to 160 basis points higher than loans backing non-GSE pools. The second column in each panel displays results from regressions that control for a full

[9] In the Online Appendix main specifications are reestimated using a logit model.

set of loan and borrower characteristics, which includes relevant underwriting variables, as well as economic factors that might affect default rates after origination, such as county-level unemployment rates and house price indices. The sign of the GSE coefficient flips, becoming negative and statistically significant. Loans in GSE pools default by approximately one percentage point less than loans in non-GSE pools, controlling for observable loan and borrower characteristics and local economic conditions.[10] This pattern suggests that the GSEs purchased subprime PLS securities composed of observably riskier mortgages but that those mortgages performed better in ways that are unobservable to the econometrician and were likely unobservable to PLS investors at the time of contracting. The third column in each panel adds deal fixed effects to the specification and thus uses only within-deal variation to estimate the difference in performance between pool types. The addition of deal fixed effects increases the absolute magnitude of the coefficient estimates by almost a factor of two. Mortgages in GSE pools default, on average, 150 to 190 basis points less than loans in non-GSE pools in the same subprime PLS deal. These results show the importance of controlling for such deal-level factors as the issuer, the originator, and subordination levels.[11]

The observation that mortgages in GSE pools perform much better than those in non-GSE pools in the same deals after controlling for a large set of observable characteristics suggests that the loans backing GSE pools are different in unobservable ways. There are a couple of potential explanations for this pattern. The first is that issuers used private information about the quality of the mortgages to give the GSEs higher-quality loans within the same deal. Since Fannie Mae and Freddie Mac were such important investors in the market, PLS issuers would have had an incentive to maintain a good reputation with the two institutions in order to ensure a stable source of demand for future business. A second

[10]In order to conserve space, the coefficient estimates associated with the borrower and mortgage characteristics included as covariates are not reported here. However, in the Online Appendix these estimates are reported for one of the regression specifications.

[11]The Online Appendix also presents results based on horizons measured from the issuance date rather than up to a fixed calendar date as in the main tables. Specifically, results are shown using 24-month and 36-month default horizons from the month of issuance, following, among others, Demyanyk and Van Hemert (2011). The point estimates are very similar to those reported in Table 3.

possible explanation is simply omitted variable bias. It could be the case that GSE pools had different risk characteristics that are not observed now but were observable to other investors in the market, and this resulted in the loans in GSE pools performing better. One such potential characteristic might be, for example, the inclusion of loans in GSE pools that were eligible for the affordable housing goals. This concern can be addressed directly by including very fine geographic fixed effects (discussed in more detail in Section 4.5 below). More generally, however, a number of cross-sectional tests and robustness tests are run to distinguish between the two explanations. The results suggest that omitted variable bias is an unlikely explanation for the main findings in the paper.

4.1 Low-documentation loans

One natural place to examine whether issuer private information could explain the results is to look specifically at low-documentation mortgages. The existing literature has argued that low-documentation lending is the segment of the market where banks putting together the deals are most likely to have private information that could lead to systematically differential performance across mortgage pools that is not accounted for by observable loan characteristics (e.g, Begley and Purnanandam (2017); Jiang et al. (2014a); Jiang et al. (2014b); Keys et al. (2012); and Saengchote (2013)).

Returning to Table 3, the fourth column of each panel shows that the difference in default risk between mortgages in GSE and non-GSE pools is significantly reduced (by 50 percent or more) once the regressions include an interaction between the GSE dummy and an indicator variable for low-documentation mortgages. (Note that a low-documentation indicator is already included in the model as an underwriting characteristic.) This means that performance differences are relatively small for full-documentation loans across GSE and non-GSE pools, but that these differences are significantly amplified in the sample of low-documentation loans. Specifically, we find that low-documentation loans in GSE pools default by 2.6 to 3.2 percentage points less on average compared to low-documentation loans in non-GSE pools. This evidence is consistent with private information on the part of subprime PLS issuers driving the differences in performance between loans in GSE pools and loans in non-GSE pools. If differences in pool performance were driven by omitted variables it is not clear why so much of the effect would be concentrated in low-documentation loans.[12]

4.2 Proportion of deals with GSE participation

To test the private information hypothesis further, we construct a measure of the frequency of interactions between the GSEs and subprime PLS issuers. Specifically, for each deal issuer in the CoreLogic dataset, the number of deals by that issuer that contain a GSE pool up to a given quarter is divided by the total number of deals issued by the issuer from the beginning of the sample until that same quarter. In the resulting variable, the "GSE Deal Fraction," a value of one means that all of the issuer's previous deals involved the GSEs, while a value of zero means that none involved the GSEs. This variable measures the extent of issuer dependency on the GSEs at a given point in time, such that an issuer with a high GSE Deal Fraction will have an incentive to supply the GSEs with higher-quality loans in order to maintain a good reputation. If the difference in performance is largely explained by loans in pools arranged by issuers with high GSE Deal Fractions, then a private information story is the most likely explanation (as it is unclear why an omitted variables explanation would be systematically related to this interaction variable).

Table 4 shows descriptive statistics of the GSE Deal Fraction variable for each year in the sample period. Some issuers almost exclusively created deals that included pools aimed at the GSEs (e.g., Fremont had a GSE pool in 100 percent of their deals, as did Wells Fargo, Barclays, and Fieldstone in 2004 and 2005). But the mean of this variable across all issuers is about 60 percent, suggesting that other issuers also sold a significant fraction of deals

[12] In the Online Appendix regressions estimated separately for each quarter (equivalent to including a full set of interactions between all of the controls and quarter of issuance indicator variables) show that the results are not driven by one or two quarters alone, but rather that the effects of both the GSE dummy and the low-documentation interaction are always below zero, and become especially strong after the first quarter of 2005.

without a pool specifically directed at the GSEs. One notable observation from the table is that there is substantial variation in the GSE Deal Fraction variable both across issuers as well as over time for a given issuer.

Table 5 assesses the extent to which prior interaction with the GSEs affects the finding that within-deal mortgage performance was better for GSE pools. An interaction between the GSE Deal Fraction variable and the GSE dummy is added to the specification in Table 3. Columns (1)–(3) in Table 5 display the estimation results for the three different default horizons, and in each case the interaction term is negative and statistically significant. Thus, a loan in a GSE pool arranged by an issuer with a GSE Deal Fraction of one is approximately 2.6 to 3.3 percentage points less likely to default than a loan in a GSE pool arranged by an issuer with no prior experience with the GSEs (i.e, a GSE Deal Fraction of zero). The interaction between the GSE dummy and the low-documentation dummy is largely unaffected, remaining negative and statistically significant. However, the addition of the new interaction term causes the sign of the GSE dummy coefficient to flip from negative to positive. This suggests that full-documentation loans in GSE pools arranged by issuers with very little or no prior experience with the GSEs are riskier than similar loans in non-GSE pools within the same deal. Thus, controlling for documentation status and previous experience with the GSEs fully explains the performance differential between mortgages in GSE pools and non-GSE pools in the same deal.

4.3 Issuer-originator affiliation

The previous section demonstrated that subprime PLS issuers with a greater proportion of deals that included securities designed for Fannie Mae and Freddie Mac delivered unobservably better-quality low-documentation mortgages into GSE pools. This is consistent with those issuers having private information and using it to sort loans into pools. A natural question is how PLS issuers might come by such information, since it is the originator rather than the issuer that directly interacts with borrowers and underwrites mortgages. Private information could be transferred from originators to issuers in a couple of ways. First, there are direct relationships between many issuers and originators in the subprime PLS market. In some cases the originator and issuer are the same institution, while in others they are part of the same vertically integrated corporation (in which case the originator is typically a subsidiary of the issuer). To the extent that issuers are more likely to obtain private information about loans that are underwritten by affiliated originators (an argument also made by Demiroglu and James (2012) and Furfine (2014)), we would expect to find stronger results for the sample of loans in which the issuer and originator are affiliated corporations. Second, it is also possible for private information to be transferred between unaffiliated originators and issuers. For example, an issuer may have sufficient experience with a group of originators to be able to identify those that are especially meticulous in their screening of loans (beyond what can be inferred from the set of observable borrower characteristics). This paper assumes, however, that it is likely easier to transfer private information when there is a direct affiliation.

Columns (4)–(9) of Table 5 replicate the first three columns in the same table, but separate deals based on whether the issuer and originator were affiliated at the time that the deal was issued (either the same institution or part of the same vertically integrated corporation). Approximately two-thirds of the observations in the CoreLogic database (64.2 percent of the sample) contain information on the identity of the originator of the mortgages included in the deal, and for this subset we are able to determine whether the originator and issuer are affiliated. The columns with issuer-affiliated originators include 396 deals where all loans are by affiliated originators, and the "unaffiliated" column has 695 deals where no loans are made by issuer-affiliated originators. This leaves out 85 deals that had a mix of both affiliated and unaffiliated issuers.[13]

We find that within issuer-originator affiliated deals (columns (4)–(6)), low-documentation

[13] The 85 mixed deals are dropped due to a lack of confidence in the identity of the originator and/or the ability to identify a relationship between the issuer and originator (the raw data on originator identities in the CoreLogic database is somewhat messy and required significant effort in cleaning and standardizing the names in order to integrate the information into the empirical analysis).

mortgages in GSE pools perform significantly better compared to those in non-GSE pools; and that the fraction of previous deals made with the GSEs is strongly correlated with performance. For unaffiliated deals, negative point estimates are obtained for both low-documentation and GSE Deal Fraction variable interactions, but the results are statistically and economically weaker (especially for the 2010 and 2012 horizons). This is consistent with better transmission and use of private information within the same organization, and suggests that unaffiliated issuers were less likely to have private information that they could use to form mortgage pools.

4.4 Estimation on a Matched Sample

Perhaps the most obvious concern regarding the results presented above is that differences in observable mortgage and borrower characteristics between loans in GSE pools and loans in non-GSE pools are not being adequately controlled for in the estimation. Table 2 (columns (1)–(4)) shows that there is covariate imbalance across the GSE and non-GSE loan pools, so that simply controlling in a linear fashion for observable characteristics may not be enough. This section addresses this concern by constructing a matched sample of loans in GSE and non-GSE pools and reestimating empirical models on this matched sample.

Because of the numerous covariates, a nearest-neighbor match based on propensity scores is conducted. The first step is estimating the propensity score, which in this context is simply the likelihood that a loan is placed in a GSE pool. Following the convention in the literature, a logit model is estimated (e.g., Caliendo and Kopeinig (2008) and Imbens and Wooldridge (2009)). The same set of covariates are included in the propensity score model as the ones used in the estimation above, with one important exception: the omission of deal fixed effects in the propensity score model (because, as explained below, loans within deals are then matched).[14] The Online Appendix shows the estimated propensity score distributions

[14]The same relatively simple specification is used for the propensity score model as is used for the linear probability model discussed above with the covariates entering the model in a linear fashion. However, because most of the covariates are specified as indicator variables, this specification constitutes a fairly rich

for loans in GSE and non-GSE pools.

The second step involves matching on the estimated propensity scores. A nearest-neighbor match is implemented, and the match is performed within deals. That is, loans in GSE pools are matched with loans in non-GSE pools only if they were included *in the same subprime PLS deal.* Matching within deals permits the use of a deal fixed effects estimator that parallels the one presented above on the full unmatched sample. We match without replacement and impose a conservative caliper of 0.02, which means that matches are kept only if the difference in their respective propensity scores is within 0.02. Columns (5) and (6) of Table 2 displays the summary statistics for the matched sample. The sample size decreases significantly (by about two-thirds compared to the full unmatched sample),[15] but the matched sample still consists of more than 3 million mortgages. It is apparent from the table that the covariate means across loans in GSE and non-GSE pools are significantly closer for the matched sample. For example, the difference in the share of adjustable-rate mortgages is 3.4 percent in the matched sample compared to 25.5 percent in the full, unmatched sample.

Although matching results in a much better covariate balance, some differences between loans in GSE and non-GSE pools remain. For this reason, regressions on the matched sample still control for differences in observable loan and borrower characteristics, as opposed to implementing a simple difference in means estimator that has been frequently used in the propensity score matching literature. Table 6 displays estimation results using the matched sample of mortgages. For each default horizon (2008, 2010, and 2012), the first column shows the within-deal difference in performance of loans in GSE pools and those in non-GSE pools (these estimates can be compared directly to the third column of each horizon in Table 3). The second column for each horizon shows the difference in performance of low-documentation loans included in GSE pools (comparable to the fourth column of Table

model. Experiments with more complicated propensity score models that include numerous interaction terms between the covariates, yield very similar results.

[15]Matching with replacement leaves a much larger sample (approximately 8.3 million loans compared to 3.4 million loans) but results in many duplicate untreated mortgages (loans in non-GSE pools). In some deals, for example, a single untreated loan was matched to thousands of treated loans. For this reason, the match was conducted without replacement.

3). The third column shows the effect of adding an interaction of the GSE dummy and the GSE Deal Fraction variable (directly comparable to the results in Table 5 for all deals).

The results in Table 6 are very consistent with those in Table 3. The difference in default rates between loans in GSE and non-GSE pools within the same deals and controlling for all relevant observable borrower and loan characteristics is between 1.5 and 1.8 percentage points (first column of each horizon panel), which is an almost identical magnitude to the result found in Table 3. Furthermore, the difference in default rates is again significantly larger in the sample of low-documentation loans. Similarly, the interaction between the GSE pool indicator and the GSE Deal Fraction variable is negative and statistically significant with very similar magnitudes to those obtained in the full, unmatched sample in Table 5. The Online Appendix also shows that the results broken down by issuer and originator affiliation are similar to those in Table 5.

4.5 Additional Robustness Tests

We view the tests using the matched sample as an important robustness test, as they largely dismiss concerns that nonlinear effects due to observable characteristics might influence the results. The results also suggest that characteristics that are unrelated to the presence of the GSEs are unlikely to be an important determinant of the results, as the magnitudes are strikingly similar to those found in the full sample. Although the matching exercise is a powerful robustness check, concerns regarding the validity and the interpretation of our results remain; these are addressed in a series of robustness that are detailed and displayed in the Online Appendix.

The robustness exercises include: (1) the estimation of all specifications with zip code fixed effects in order to mitigate the influence of time-invariant, unobserved geographic differences; (2) the estimation of all specifications on the sample of low-documentation mortgages only to address the concern that such loans might be differentially sensitive to observable risk characteristics; (3) the exclusion of jumbo loans from the estimation (i.e, only loans below the CLL); (4) the exclusion of fixed-rate mortgages from the sample in order to mitigate potential bias driven by differences in exposure to prepayment risk across GSE and non-GSE pools; (5) the inclusion of controls for loans that qualify for GSE affordable housing goals credit; and (6) the exclusion of deals with multiple servicers so that the deal fixed effects fully account for servicer heterogeneity.

In addition, the results are robust to changing the default definition to 90+ days delinquent, to changing the default horizon to 24 and 36 months after origination, to assuming a logit specification instead of a linear probability model, and to clustering standard errors by deal instead of by quarter of issuance.

4.6 Comparison of ex-ante risk characteristics

This section compares the riskiness of the mortgages underlying GSE and non-GSE pools based on only the borrower and loan characteristics available when the deals were issued. A comparison of ex-ante default risk permits the verification of whether differences in loan performance between GSE and non-GSE pools were predictable by investors using only information at the time of contracting.

Ex-ante default probabilities are constructed for each mortgage in the sample in the spirit of the model in Ashcraft et al. (2010). The quarter in which the corresponding deal was issued is determined for each loan in the sample. All loans in deals issued between 24 months and 12 months *before* that quarter are then taken and tracked over the subsequent 12 months, creating indicator variables that take a value of one if the mortgage is 60 days delinquent, 90 days delinquent, in foreclosure, or in REO (or any other liquidation status following foreclosure) at any point during the 12-month period, respectively.[16] The same exercise is performed for horizons of 24 and 36 months.[17] Three discrete choice models are

[16]The 12-month predicted default variable uses loans originated between 24 months and 12 months before that quarter, so that each loan has a full 12 months of history.

[17]For the 24-month horizon, all loans that collateralized deals issued between 36 months and 24 months before the quarter of interest are tracked over the subsequent 24 months, while for the 36-month horizon, all loans in deals issued between 48 and 36 months before the quarter of interest are tracked over the subsequent

then estimated: a linear probability model, a logistic regression, and a multinomial logistic regression that specifically accounts for the fact that mortgages can prepay as well as default. In addition to these three models, a variant of a competing risks hazard model is estimated using a multinomial logit specification.

The regressions are estimated each quarter over the period 2003-2007 and include the same borrower and loan characteristics reported in Table 2 as well as state fixed effects. The estimated coefficients from these loan-level credit risk models are applied to the characteristics of the loans in deals issued in the current quarter to create the 12-month, 24-month, and 36-month loan-level default probabilities. This means that ex-ante default probabilities are created using only information available when the deals are issued.

Table 7 shows the results from loan-level regressions of the 12-month, 24-month, and 36-month ex-ante (predicted) default probabilities on a dummy variable that is equal to one for GSE pools. As discussed above, four alternative models are used to compute the default probabilities at the loan level. Panels A, B, C, and D correspond to the linear probability model, logit model, multinomial logit model, and competing risks hazard model, respectively.

The results in Table 7 show that loans in GSE pools either do not look riskier or the difference in predicted default probabilities is, in fact, positive, meaning that the GSE pools look somewhat riskier using information available at origination. The linear probability model shows statistically insignificant and economically very small differences in the ex-ante riskiness of loans in GSE pools versus those in pools directed at other investors, whereas the logit and multinomial logit models suggest that there is a higher ex-ante default probability of 60-100 basis points at a two- or three-year horizon (although there is no difference at the 12-month horizon). Note that a very large share of the variation in ex-ante mortgage risk can be explained by the deal fixed effects included in these regressions (R^2 goes from 1 percent to over 30 percent). These results confirm that there were no observable differences between GSE and non-GSE pools, and that the differences that emerge ex-post are likely the result

36 months.

of *unobservable* differences between pools.

4.7 Analysis of PLS yields

Thus far this paper has focused on differences in loan quality between GSE and non-GSE PLS pools and has documented the fact that the GSEs invested in pools characterized by loans that performed significantly better than those in non-GSE pools. These quality differences are interpreted here to have resulted from preferential treatment by subprime PLS issuers who sorted loans using private information that was unavailable to other PLS investors. An additional, and perhaps more straightforward, dimension on which the GSEs may have received preferential treatment is pricing. This section explores whether there is also evidence of a GSE catering effect in the pricing of PLS securities.

Two issues need to be addressed before the details and results of the analysis are presented. First, security prices are not directly observed, and thus, the analysis instead focuses on yield spreads (specifically, the yield spread over the one-month LIBOR). We do not believe that this is an important drawback in the context of this study, because the analysis covers only floating-rate tranches and inverse floaters. This minimizes the role of interest rate risk and the negative convexity issue that arises with fixed-rate, long-duration, mortgage-backed securities. Crucially, this means that the vast majority of the securities were issued at par, and thus variation in yield spreads will be very highly correlated with variation in prices (at issuance).

The second issue relates to the analysis of performance differences above. The alternative explanation of greatest concern is that the performance differences between loans in GSE and non-GSE pools are due to differences in the characteristics of the loans that are observable to PLS investors and issuers but that are not observable to us. If this interpretation is correct and the performance differences were forecastable by PLS investors, then they should be priced into the securities. Specifically, all else equal, higher prices (lower yields) should be associated with the triple-A GSE securities compared to the triple-A non-GSE securities within the same subprime deal. Thus, this analysis of yield spreads is both an interesting exercise in its own right and an additional test of the hypothesis that the performance differences between GSE and non-GSE mortgage pools are caused by omitted variable bias.[18]

To operationalize this exercise, this section focuses on the yield spreads of triple-A securities with claims on GSE and non-GSE pools and limits the sample to deals in which all the triple-A securities were either floating-rate tranches or inverse floaters. Focusing on floating-rate tranches has two main advantages: first, because the yield on those tranches is always quoted as a spread over one-month LIBOR in our sample, the yield can be cleanly aggregated for multiple tranches in the same deal and construct a pool-level spread. Second, because these tranches have a very short duration,the prepayment risk that arises with fixed-rate, long-duration, mortgage-backed securities can be ignored. Including only the floating-rate mortgage pools drops 177 observations, leaving 3,290 unique pools for this set of regressions. The Online Appendix displays summary statistics (from Bloomberg) for the triple-A securities that are collateralized by the loans in the CoreLogic database broken down by whether the securities are derived from GSE or non-GSE mortgage pools.

Table 8 reports the results of regressions of the at-issuance yield spreads of the securities in the sample (relative to LIBOR) on the GSE pool indicator variable, a control for the weighted average of the expected life of each security, and average pool-level loan characteristics. There is no evidence that the yield spreads associated with GSE triple-A securities were lower than corresponding non-GSE triple-A securities. Instead, in fact, the yield spreads on GSE-purchased securities are, on average, 3–6 basis points *higher* than those purchased by non-GSE investors. This holds both with and without deal fixed effects, and also when controlling for observable characteristics of the mortgages in each pool.[19] This suggests

[18]Other studies in the literature that have analyzed PLS yield spreads include Demiroglu and James (2012), which finds that average yield spreads are lower for deals in which the issuer and originator are affiliated, and He et al. (2012), which finds that deals sold by large issuers have higher yields than deals sold by small issuers.

[19]Specifically, the regressions include pool-level average mortgage characteristics. A potential concern is that if non-linearities are important, omitting higher-order moments of the distribution may affect the results. Experimentx with specifying FICO scores in this way produced results that were largely unaffected.

that Fannie Mae and Freddie Mac were able to obtain higher-yielding securities relative to other investors in the same deals buying similarly triple-A rated securities. These are large magnitudes relative to the average spread over LIBOR, which ranged from 13 to 39 basis points over the sample period. Columns (3) and (6) illustrate that much of the difference in spreads at origination can be explained by the interaction of the GSE dummy with GSE Deal Fraction, suggesting that Fannie Mae and Freddie Mac received particularly good deals from issuers that frequently included GSE pools in their deals.

The results displayed in Table 8 are consistent with a GSE catering effect on the price dimension in addition to the quality dimension documented above. Furthermore, the fact that the GSEs appear to have received higher yields (lower prices) compared to the other PLS investors in the market also supports the contention that the performance results were not forecastable by investors, but rather, likely driven by private information unknown by other PLS investors.

4.8 Across-deal comparison of non-GSE loan pools

The bulk of the analysis thus far has used within-deal variation between loans in GSE and non-GSE pools. This section exploits across-deal variation to shed light on whether investors in deals with a GSE presence were harmed relative to those that stayed away from deals where the GSEs participated. A simple exercise is implemented that involves comparing the performance of loans in non-GSE pools in deals *with* GSE participation to those in deals *without* GSE participation. Given that this is an across-deal analysis, fixed effects cannot be included.

Panel A of Table 9 displays the results. The variable of interest is "GSE Deal," an indicator variable that distinguishes between non-GSE loans in PLS deals with and without GSE participation. Column (1) shows that loans in deals with GSE participation defaulted by 2.6 percentage points more than loans in deals without GSE participation. Column (2) shows that this difference is even larger among the sample of low-documentation loans (3.6

percentage points), and column (3) shows that the performance difference is especially large in deals where the issuer had a strong business relationship with the GSEs. These results are consistent with the findings from the within-deal analysis of GSE versus non-GSE mortgages.

This comparison, while informative, does potentially suffer from endogenous matching issues. There may be unobserved factors that determine participation in GSE versus non-GSE deals, which also explain the differences in average loan performance. To partially address this issue, columns (4)–(6) include, in addition to the detailed set of mortgage and borrower controls, a full set of issuer fixed effects. The results are largely unchanged, with the exception of the coefficient of the "GSE Deal * Deal Fraction" interaction term, which is reduced.

Finally, Panel B of Table 9 compares the average yield spreads associated with PLS triple-A securities purchased by non-GSE investors in deals with and without GSE participation. The results show that yield spreads were lower in deals that included the GSEs, again consistent with worse outcomes for investors that participated in deals where the GSEs were also present. Overall, the results from this analysis are consistent with the idea that the GSE catering effect identified above may have had a negative affect on other PLS investors in the market. However, due to some endogeneity issues that cannot be fully addressed in the analysis (specifically the inability to include deal fixed effects), a more definitive answer to this question requires future research.

5 Conclusion

Fannie Mae and Freddie Mac have long played a central role in the U.S. housing finance system as both securitizers and investors in conforming prime mortgages. Moreover, during the recent U.S. housing boom, the GSEs were also the two largest investors in subprime PLS– a fact that has largely escaped academic attention. This paper uses a unique feature of the structure of subprime PLS deals to show that, conditional on observable risk characteristics, the loan pools that were eligible to be bought by Fannie Mae and Freddie Mac performed significantly better during the crisis relative to mortgage pools backing securities sold to other investors in the same deals. This difference is concentrated in low-documentation loans, which suggests that issuers were using private information to sort mortgages into GSE and non-GSE pools. Deals sold by issuers that frequently structured deals for Fannie Mae and Freddie Mac also exhibit larger differences in performance between GSE and non-GSE pools, especially when the issuer of the deals and the originator of the mortgages are affiliated institutions. In addition, the GSEs appear to have received more favorable pricing compared to other PLS investors. These findings support the view that subprime PLS issuers gave beneficial treatment to the GSEs in these markets. This was likely due to reputational concerns and/or the perception that the GSEs, because of their size, would be more likely to trigger contractual clauses in the event of wide-scale mortgage defaults. Of course, this scenario came to fruition in the aftermath of the global financial crisis.

The results are consistent with the view that the presence of large, concentrated claimholders can influence security design and performance. This is especially striking for two reasons. First, the GSEs have been accused of having heavily distorted incentives that led them to take excessive risks. This means that, if anything, this study's tests are biased against finding any risk-mitigating effect by the GSEs. Second, the securities purchased by the GSEs were largely considered to be information-insensitive at the time of issuance, as they were rated triple-A. This means that the role of large and concentrated investors can influence security design and performance well beyond corporate equity and bank debt, where their role has been well established by the literature.

References

Ashcraft, A., P. Goldsmith-Pinkham, and J. Vickery (2010). MBS ratings and the mortgage credit boom. Staff report #449, Federal Reserve Bank of New York.

Begley, T. A. and A. Purnanandam (2017). Design of financial securities: Empirical evidence from private-label RMBS deals. *Review of Financial Studies 30*(1), 120–161.

Boot, A. W. (2000). Relationship banking: What do we know? *Journal of Financial Intermediation 9*(1), 7–25.

Brav, A., W. Jiang, F. Partnoy, and R. Thomas (2008). Hedge fund activism, corporate governance, and firm performance. *The Journal of Finance 63*(4), 1729–1775.

Bunce, H. L. (2007). *The GSEs' Funding of Affordable Loans: A 2004-05 Update*. Office of Policy Development and Research, U.S. Department of Housing and Urban Development.

Burkart, M., D. Gromb, and F. Panunzi (1997). Large shareholders, monitoring, and the value of the firm. *The Quarterly Journal of Economics 112*(3), 693–728.

Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys 22*(1), 31–72.

Degryse, H. and P. Van Cayseele (2000). Relationship lending within a bank-based system: Evidence from European small business data. *Journal of Financial Intermediation 9*(1), 90–109.

Demiroglu, C. and C. James (2012). How important is having skin in the game? Originator-sponsor affiliation and losses on mortgage-backed securities. *Review of Financial Studies 25*(11), 3217–3258.

Demyanyk, Y. and O. Van Hemert (2011). Understanding the subprime mortgage crisis. *Review of Financial Studies 24*(6), 1848–1880.

Edmans, A. (2009). Blockholder trading, market efficiency, and managerial myopia. *The Journal of Finance 64*(6), 2481–2513.

Ellul, A., C. Jotikasthira, C. T. Lundblad, and Y. Wang (2014). Mark-to-market accounting and systemic risk: Evidence from the insurance industry. *Economic Policy 29*(78), 297–341.

FCIC (2011). *The financial crisis inquiry report: final report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*. Government Printing Office.

Furfine, C. H. (2014). Complexity and loan performance: Evidence from the securitization of commercial mortgages. *Review of Corporate Finance Studies 2*(2), 154–187.

Ghent, A. C., R. Hernández-Murillo, and M. T. Owyang (2015). Did affordable housing legislation contribute to the subprime securities boom? *Real Estate Economics 43*(4), 820–854.

Greenlaw, D., J. Hatzius, A. K. Kashyap, and H. S. Shin (2008). Leveraged losses: Lessons from the mortgage market meltdown. In *Proceedings of the US Monetary Policy Forum*, pp. 7–59.

Greenspan, A. (2005). Regulatory reform of the government-sponsored enterprises: testimony before the committee on banking, housing, and urban affairs, us senate, april 6, 2005. Technical report, Board of Governors of the Federal Reserve System (US).

He, J. J., J. Q. Qian, and P. E. Strahan (2012). Are all ratings created equal? The impact of issuer size on the pricing of mortgage-backed securities. *The Journal of Finance 67*(6), 2097–2137.

Imbens, G. W. and J. M. Wooldridge (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature 47*(1), 5–86.

Jiang, W., A. A. Nelson, and E. Vytlacil (2014a). Liar's loan? Effects of origination channel and information falsification on mortgage delinquency. *Review of Economics and Statistics 96*(1), 1–18.

Jiang, W., A. A. Nelson, and E. Vytlacil (2014b). Securitization and loan performance: Ex ante and ex post relations in the mortgage market. *Review of Financial Studies 27*(2), 454–483.

Keys, B. J., A. Seru, and V. Vig (2012). Lender screening and the role of securitization: Evidence from prime and subprime mortgage markets. *Review of Financial Studies 25*(7), 2071–2108.

McCahery, J. A., Z. Sautner, and L. T. Starks (2016). Behind the scenes: The corporate governance preferences of institutional investors. *The Journal of Finance 71*(6), 2905–2932.

Merrill, C. B., T. Nadauld, R. M. Stulz, and S. M. Sherlund (2014). Were there fire sales in the RMBS market? Working paper, Ohio State University, Charles A. Dice Center for Research in Financial Economics.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica: Journal of the Econometric Society 16*(1), 1–32.

Parrino, R., R. W. Sias, and L. T. Starks (2003). Voting with their feet: Institutional ownership changes around forced CEO turnover. *Journal of Financial Economics 68*(1), 3–46.

Saengchote, K. (2013). Soft information in the subprime mortgage market. Unpublished manuscript, Northwestern Kellogg School of Management.

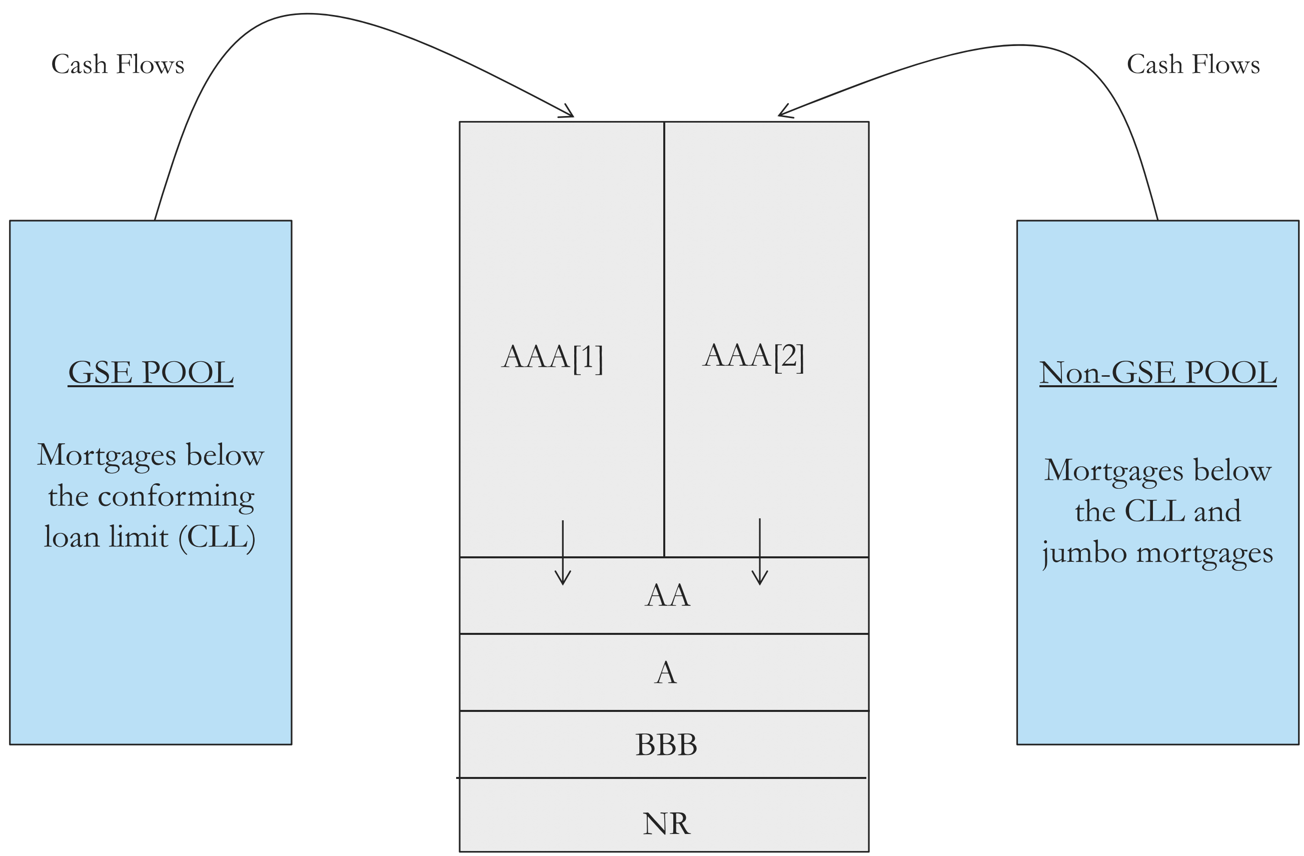
Schenone, C. (2010). Lending relationships and information rents: Do banks exploit their information advantages? *Review of Financial Studies 23*(3), 1149–1199.

Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The Journal of Finance 45*(4), 1069–1087.

Shleifer, A. and R. W. Vishny (1986). Large shareholders and corporate control. *Journal of Political Economy 94*(3, Part 1), 461–488.

Shleifer, A. and R. W. Vishny (1997). A survey of corporate governance. *The Journal of Finance 52*(2), 737–783.

Figure 1: Typical Subprime PLS Deal Structure with GSE Participation



Notes: This figure displays the structure of a typical subprime PLS deal purchased by the GSEs. These deals involved more than one mortgage pool: one consisting of only mortgages below the CLL ("GSE pool") and at least one other pool made up of mortgages below the CLL and jumbo mortgages ("Non-GSE Pool"). The lower rated securities derived their cash flows from all pools, while the triple-A securities purchased by the GSEs derived their cash flows exclusively from the GSE pool and the triple-A securities purchased by other investors derived their cash flows from the other pools.

Table 1: Subprime Private-label Securities Issuance (PLS): 2000–2008

	Subprime PLS Issuance		GSE Subprime PLS Purchases			
	# Deals	Year	FHFA Report to Congress	CoreLogic Data	Market Share	Fraction of Deals Participated In
	(1)	(2)	(3)	(4)	(5)	(6)
2000	74	52.5	.	.	.	0
2001	107	87.1	.	3.4	3.8%	0.364
2002	157	122.7	.	14.6	11.9%	0.516
2003	238	195.0	.	67.7	34.7%	0.660
2004	350	362.6	.	141.0	38.9%	0.720
2005	457	465.0	.	134.4	28.9%	0.626
2006	523	448.6	110.4	106.0	23.6%	0.547
2007	241	201.6	59.6	50.1	24.9%	0.635
2008	0	2.3	0.7	0	.	.

Notes: Subprime PLS Issuance is obtained from the 2011 Mortgage Market Statistical Annual (volume II, page 31). Publicly available data on PLS purchased by the GSEs is obtained from the 2012 Federal Housing Finance Agency's (FHFA) Annual Report to Congress. The FHFA report only breaks out PLS purchases into subprime and Alt-A for Freddie Mac beginning in 2006. Proprietary data on PLS purchased by the GSEs is obtained from CoreLogic's Asset-Backed Securities database. The GSEs' market share of PLS purchases (column 6) is obtained by dividing GSE PLS purchases (column 4) by total subprime PLS issuance (column 3). The number of subprime PLS deals (column 2) and the fraction of deals in which the GSEs' purchased securities is calculated from the CoreLogic database.

Table 2: Loan-level Summary Statistics: Corelogic Subprime PLS Issued 2003–2007

	Full sample (unconditional)		Full sample (conditional on deal F.E.)		Matched Sample	
	Non-GSE	GSE	Non-GSE	GSE	Non-GSE	GSE
	N = 6,324,311	N = 4,140,711	N = 6,324,311	N = 4,140,711	N = 1,724,149	N = 1,724,149
Continuous Variables	Mean	Mean	Mean	Mean	Mean	Mean
	(1)	(2)	(3)	(4)	(5)	(6)
FICO (Points)	642	616	635	626	618	624
Balance ($)	159,224	156,907	183,399	119,987	163,977	156,558
CLTV (P.Points)	88.8	84.4	87.8	85.8	86.5	85.9
Orig. Rate (P. Points)	8.64	7.94	8.40	8.31	8.22	8.10
Term (months)	314	350	327	332	349	346
Unemployment (P. Points)	5.09	5.39	5.18	5.25	5.22	5.15
Trailing 12-month unemployment change	-6.5%	-5.3%	-6.2%	-5.8%	-6.3%	-6.1%
Unemployment change through 2012	54.7%	47.1%	63.1%	59.7%	62.7%	61.7%
Price Index	181	168	179	170	177	174
Trailing 12-month HPA	12.1%	12.3%	12.5%	11.6%	12.0%	11.9%
HPA through 2012	-17.5%	-13.8%	-16.4%	-15.6%	-17.0%	-16.5%
Indicator Variables						
Low Documentation (share)	0.412	0.347	0.399	0.361	0.365	0.342
Non-Owner Occupied (share)	0.083	0.084	0.077	0.095	0.072	0.080
Purchase Loan (share)	0.508	0.356	0.499	0.366	0.495	0.439
Cash-Out Refinance (share)	0.422	0.563	0.432	0.548	0.442	0.482
Interest-Only (share)	0.137	0.096	0.138	0.090	0.114	0.122
Balloon (share)	0.225	0.094	0.175	0.167	0.134	0.126
ARM (share)	0.489	0.744	0.549	0.650	0.688	0.654
Prepay Penalty (share)	0.616	0.719	0.646	0.631	0.733	0.699
Default Rate through 2008:Q4	0.338	0.315	0.342	0.308	0.385	0.344
Default Rate through 2010:Q4	0.421	0.376	0.416	0.383	0.458	0.424
Default Rate through 2012:Q4	0.444	0.393	0.436	0.405	0.477	0.447

Notes: This table shows summary statistics for both the full sample of loans underlying the GSE and non-GSE triple-A subprime PLS securities in the Corelogic database issued between 2003 and 2007 as well as in a matched sample of loans in GSE and non-GSE loan pools. The details of the construction of the matched sample are discussed in section 4.2 of the text. The table displays both unconditional full sample means and full sample means conditional on deal fixed effects. GSE refers to mortgage pools made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. The variables coincide with the covariate set included in the ex-ante and ex-post default regressions below. FICO score is the credit score of the borrowers on the loan application; balance is the size of the loan at origination; CLTV is the size of all loans on the property relative to the price of the house (transaction price or appraisal amount, depending on whether it is a purchase or a refinance); original rate is the interest rate on the loan at origination; term is the original term of the loan; unemployment is measured at the county-level (from the BLS); HPA is the county-level house price appreciation (from CoreLogic); low-documentation is a 1 if the loan was either low-documentation or no documentation; non-owner occupant is a 1 if the property is for investment purposes or is a vacation/second home. The bottom of the table shows the average realized default rates of each loan type through 2008:Q4, 2010:Q4, and 2012:Q4.

Table 3: Effect of GSE participation on Default

	Horizon through 2008:Q4				Horizon through 2010:Q4				Horizon through 2012:Q4			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
GSE (d)	0.016***	-0.011***	-0.019***	-0.007***	0.014**	-0.008***	-0.016***	-0.006**	0.012**	-0.008***	-0.015***	-0.006
	(3.04)	(5.97)	(10.69)	(4.17)	(2.49)	(4.03)	(8.77)	(2.53)	(2.30)	(3.74)	(7.99)	(2.55)
Low Doc			0.057***	0.070***			0.060***	0.072***			0.058***	0.069***
			(8.22)	(10.04)			(8.94)	(10.77)			(9.58)	(11.60)
GSE*Low Doc				-0.032***				-0.029***				-0.026***
				(9.19)				(8.76)				(8.73)
Deal F.E. ?	N	N	Y	Y	N	N	Y	Y	N	N	Y	Y
Covariates ?	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
Issue Year F.E. ?	Y	Y	.	.	Y	Y	.	.	Y	Y	.	.
# Loans	10,465,022	10,465,022	10,464,165	10,464,165	10,465,022	10,465,022	10,464,165	10,464,165	10,465,022	10,465,022	10,464,165	10,464,165
# Deals	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809
Adjusted R^2	0.04	0.14	0.16	0.16	0.09	0.19	0.20	0.20	0.11	0.20	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 4: List of Sponsors with Highest Values of "GSE Deal Fraction"

Sponsor	Average value of GSE Deal Fraction (%)						# Deals
	All Years	2003	2004	2005	2006	2007	(2003 - 2007)
Fremont	100	100	100	100	100	.	28
Fieldstone	98.3	.	100	100	94.1	91.7	13
Wells Fargo	94.4	.	100	100	87.4	63.6	11
Barclays	91.8	.	100	100	88.8	84.0	36
Washington Mutual	83.7	84.7	78.2	82.5	86.1	87.9	43
UBS	82.5	100	97.3	89.4	68.7	61.3	42
Morgan Stanley	80.4	75.3	79.5	83.5	81.8	78.6	111
National City	77.8	73.1	77.3	78.2	79.4	.	65
Goldman Sachs	77.3	100	91.3	78.0	70.9	69.0	65
Deutsche Bank	75.7	64.4	81.7	78.9	74.3	73.2	74
All Sponsors	59.7	43.1	59.8	62.4	61.4	58.1	1,751

Notes: This table displays the ten subprime PLS sponsors with the highest values of the "GSE Deal Fraction" measure used in the analysis. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Only sponsors involved in at least ten deals over the sample period (2003–2007) are included in the table.

Table 5: Default, "GSE Deal Fraction," and Issuer/Originator Affiliation

		All Deals			Affiliated Deals			Unaffiliated Deals	
Horizon	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GSE (d)	0.013***	0.012***	0.011***	0.026**	0.025***	0.024***	-0.005	-0.014**	-0.015**
	(3.59)	(4.03)	(4.21)	(2.43)	(2.91)	(3.01)	(0.78)	(2.24)	(2.45)
GSE*Low Doc	-0.032***	-0.029***	-0.026***	-0.034***	-0.025***	-0.022***	-0.032***	-0.009	-0.008
	(8.99)	(8.76)	(8.72)	(6.29)	(6.26)	(6.19)	(7.08)	(0.80)	(0.80)
GSE*"GSE Deal Fraction"	-0.033***	-0.027***	-0.026***	-0.052***	-0.051***	-0.050***	-0.005	0.009	0.010
	(5.73)	(4.56)	(4.43)	(3.21)	(3.58)	(3.70)	(0.58)	(1.00)	(1.14)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,156,202	10,156,202	10,156,202	2,668,773	2,668,773	2,668,773	3,374,320	3,374,320	3,374,320
# Deals	1,724	1,724	1,724	396	396	396	695	695	695
Adjusted R^2	0.16	0.20	0.21	0.15	0.19	0.21	0.15	0.19	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The first three columns display results for the sample of all deals, while columns 4-6 display results where the originator of all loans in a deal is affiliated with the issuer, while the last three columns display results for the sample of deals in which the originator is not affiliated with the issuer. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 6: Effect of GSE Participation on Default: Matched Sample

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GSE (d)	-0.018***	-0.011***	0.008***	-0.016***	-0.011***	0.005*	-0.015***	-0.011***	0.005
	(8.05)	(5.43)	(2.35)	(6.19)	(4.93)	(1.66)	(5.78)	(4.72)	(1.60)
GSE * Low Doc		-0.020***	-0.019***		-0.014***	-0.014***		-0.012***	-0.012***
		(9.99)	(9.34)		(8.52)	(8.39)		(7.44)	(7.33)
GSE*"GSE Deal Fraction"			-0.031***			-0.026***			-0.024***
			(5.39)			(4.34)			(4.34)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	3,448,298	3,448,298	3,448,298	3,448,298	3,448,298	3,448,298	3,448,298	3,448,298	3,448,298
# Deals	1,130	1,130	1,130	1,130	1,130	1,130	1,130	1,130	1,130
Adjusted R^2	0.05	0.13	0.14	0.10	0.18	0.19	0.11	0.19	0.20

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on a matched sample of loans in GSE and non-GSE pools. The details of the construction of the matched sample are discussed in section 4.2 of the text. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 7: Ex-Ante Default Probabilities for Loans in GSE and Non-GSE Pools

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0053*	-0.0064***	0.0200***	-0.0025	0.0251***	-0.0029
	(1.93)	(3.79)	(4.47)	(1.07)	(5.40)	(1.21)
Deal F.E. ?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.37	0.01	0.31	0.01	0.30

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0035	-0.0021	0.0222***	0.0063***	0.0322***	0.0071***
	(1.20)	(1.25)	(5.55)	(3.30)	(8.72)	(3.44)
Deal F.E. ?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.33	0.01	0.24	0.02	0.19

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0042	-0.0016	0.0214***	0.0077***	0.0315***	0.0103***
	(1.48)	(1.01)	(5.03)	(3.92)	(9.20)	(5.15)
Deal F.E. ?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.32	0.01	0.24	0.01	0.19

Panel D: Competing Risks Duration Model

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0019	-0.0019	0.0116***	0.007***	0.0084*	0.0106***
	(0.67)	(1.15)	(3.06)	(3.91)	(1.92)	(5.66)
Deal F.E. ?	N	Y	N	Y	N	Y
# Loans	10,344,000	10,344,000	10,344,000	10,344,000	10,344,000	10,344,000
# Deals	1,804	1,804	1,804	1,804	1,804	1,804
Pseudo R^2	0.00	0.29	0.01	0.30	0.01	0.30

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month ex-ante default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month ex-ante rates. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. In Panel A we compute ex-ante default rates using OLS regressions, while Panel B uses ex-ante default rates using logistic regressions, and Panel C uses ex-ante default rates using multinomial logistic regressions. In Panel D we compute ex-ante default probabilities using a competing risks duration model, where we assume a logistic form for the hazards. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 8: Yield Spreads for GSE and Non-GSE Pools

	(1)	(2)	(3)	(4)	(5)	(6)
GSE (d)	2.71***	3.55***	-2.41	4.68***	6.36***	2.75
	(4.59)	(3.99)	(1.28)	(7.20)	(3.88)	(1.26)
Average Life	5.41***	6.99***	6.99***	4.52***	5.49***	5.50***
	(6.53)	(6.66)	(7.66)	(4.73)	(4.78)	
GSE * "GSE Deal Fraction"			9.61***			6.57**
			(3.29)			(2.03)
Pool Characteristics?	N	N	N	Y	Y	Y
Issue Quarter FE?	Y	.	.	Y	.	.
Deal FE?	N	Y	Y	N	Y	Y
# Pools	3,290	3,290	3,290	3,290	3,290	3,290
Adjusted R^2	0.56	0.79	0.79	0.62	0.84	0.84

Notes: This table shows pool-level, OLS regressions where the dependent variable is the pool-level average spread (in percentage points) over the contemporaneous one-month LIBOR rate. The average spread is calculated by weighting the spread on individual tranches included in each pool by their original dollar amount. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with "Pool Characteristics" (columns 4,5, and 6) have pool-level controls for the loan characteristics in the pool. A full list of those controls is given in the text in Section 3. Pool-level average life is the average weighted expected life for the tranches in each pool as advertised in the prospectus where the average is weighted by the size of each tranche. The sample includes only triple-A floating-rate tranches that are part of deals where all the triple-A tranches are either floating-rate or inverse floaters. Standard errors are heteroskedasticity-robust and clustered at the quarter of origination level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table 9: Comparison of Non-GSE Loans in GSE and Non-GSE Pools

Panel A: Differences in Default Propensities

	(1)	(2)	(3)	(4)	(5)	(6)
GSE Deal	0.026***	0.022***	-0.012	0.029***	0.024***	0.007
	(4.49)	(4.26)	(1.52)	(6.14)	(5.54)	0.58
GSE Deal * Low Doc		0.014***	0.013***		0.013***	0.012***
		(3.57)	(3.61)		(2.92)	(2.86)
GSE Deal * Deal Fraction			0.065***			0.030
			(4.51)			(1.63)
Covariates ?	Y	Y	Y	Y	Y	Y
Issuer F.E. ?	N	N	N	Y	Y	Y
# Loans	6,209,878	6,209,878	6,132,891	6,209,878	6,209,878	6,132,891
Adjusted R^2	0.16	0.16	0.17	0.17	0.17	0.17

Panel B: Differences in Yield Spreads

	(1)	(2)
GSE Deal	-2.22***	-1.14*
	(3.26)	(1.87)
Average Life	6.95***	5.45***
	(7.02)	(4.22)
Pool Characteristics	N	Y
Issue Quarter F.E. ?	Y	Y
# Pools	2,020	2,020
Adjusted R^2	0.55	0.63

Notes: This table displays results from a sample of mortgages in non-GSE, PLS pool only. The independent variable of interest is "GSE Deal" which is a 0/1 indicator variable that identifies loans in non-GSE mortgage pools that are in PLS deals with GSE participation. Panel A displays OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated through 2008:Q4. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. All regressions include controls for a large number of borrower and loan characteristics including state and issue-year fixed effects (a full list is given in Section 2.1). Columns (4) - (6) include issuer fixed effects. Panel B displays results from pool-level, OLS regressions where the dependent variable is the pool-level average spread (in percentage points) over the contemporaneous one-month LIBOR rate. The average spread is calculated by weighting the spread on individual tranches included in each pool by their original dollar amount. Regressions with "Pool Characteristics" (columns 2) have pool-level controls for the loan characteristics in the pool. A full list of those controls is given in the text in Section 3. Pool-level average life is the average weighted expected life for the tranches in each pool as advertised in the prospectus where the average is weighted by the size of each tranche. The sample includes only triple-A floating-rate tranches that are part of deals where all the triple-A tranches are either floating-rate or inverse floaters. Standard errors are heteroskedasticity-robust and clustered at the quarter of origination level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Online Appendix (NOT FOR PUBLICATION)

A.1 Algorithm for Identifying GSE Pools

This section describes our algorithm for identifying pools of mortgages backing subprime PLS deals involving Fannie Mae and Freddie Mac. As described in the text, we use a unique feature of the subprime PLS market to indirectly identify triple-A subprime PLS purchased by the GSEs. PLS deals involving Fannie Mae and Freddie Mac were split into conforming and nonconforming mortgage pools. This split was necessary to facilitate the GSEs' purchases of PLS since, by law, Fannie Mae and Freddie Mac are only allowed to securitize or invest in mortgages below the conforming loan limit. The GSEs would thus purchase triple-A securities that were backed by loans from the pool(s) of exclusively conforming mortgages, while other investors would purchase securities from the pool(s) that contained both conforming and non-conforming loans.

Based on this institutional feature of the PLS market, we design an algorithm to identify mortgage pools that backed securities purchased by either Fannie Mae or Freddie Mac between 2003 and 2007.[20] The algorithm is quite simple, with the following two conditions required for a pool to be categorized as a "GSE pool":

1. At least 99 percent of loans in the pool must have a principal balance below the conforming loan limit at the time that securities in the deal are issued to investors.

2. Less than 75 percent of loans in the pool are second liens.

The first restriction is the most important. It is based on whether the loan lies above or below the conforming loan limit at the time that the deal is issued, rather than the time

[20] Unfortunately this algorithm does not allow us to distinguish PLS purchased by Fannie Mae versus Freddie Mac.

that the mortgage is originated.[21] Table A.1 below shows the conforming loan limits for the period 2000–2007, which applied to all states within the continental U.S.

Table A.1: Loan-level Summary Statistics: Corelogic Subprime Low Documentation PLS Issued 2003–2007

Year	Conforming Loan Limit (Single Family Property)
2000	\$252,700
2001	\$275,000
2002	\$300,700
2003	\$322,700
2004	\$333,700
2005	\$359,650
2006	\$417,000
2007	\$417,000
2008	\$417,000

We allow up to 1 percent of the loan pool to be composed of non-conforming mortgages to take into account potential measurement error in the data. Specifically we are concerned with potential error stemming from two variables. First, there may be cases in which the outstanding balance reported in the CoreLogic database is incorrect. Second, there may be cases in which the variable that indicates whether a property is single-family or 2-4 family is incorrectly reported. The conforming loan limits for 2-4 family properties were significantly higher than those for single-family properties. Thus, if an observation is incorrectly categorized as pertaining to a single-family home instead of a 2-4 family property, then we would likely misclassify the observation as a non-conforming mortgage.[22]

We impose the restriction on the proportion of second lien mortgages because the vast ma-

[21]This distinction is potentially important because there are often seasoned loans in the mortgage pools so that the loan amount at origination can be higher than the outstanding balance at the time that the deal is issued.

[22]Note that our algorithm only considers single-family mortgages due to the fact that 2-4 family properties were subject to different conforming loan limits depending on the exact number of units, and CoreLogic does not distinguish between properties by the number of units (i.e. it groups 2, 3, and 4 family units together).

jority of them have outstanding balances below the conforming loan limit.[23] Hence, the conforming loan limit tells us very little about whether or not the GSEs purchased securities collateralized by those loan pools.[24]

A.2 Additional Validation of Algorithm

Table 1 in the paper compares annual, aggregate GSE purchases of subprime PLS as calculated using our algorithm with those listed in the 2011 FHFA Annual Report to Congress. In the table we were only able to compare numbers for 2006 and 2007 because the FHFA report does not break out Freddie Mac's purchases by the type of security (subprime versus Alt-A), whereas it does for Fannie Mae going back to 2003. Another source of information about the aggregate purchases of subprime securities by the GSEs is the Federal Crisis Inquiry Commission (FCIC) Report (2011). In order to infer the annual purchases in 2003-2005 by Freddie Mac, we use a Figure in the FCIC report entitled "Buyers of Non-GSE Mortgage-Backed Securities" (see page 124 of the report). Of course, we cannot obtain precise numbers for Freddie Mac from the figure, but we are able to obtain approximate numbers. Table A.2 displays Fannie Mae's numbers from the FHFA Annual Report, as well as Freddie Mac's inferred numbers from the FCIC figure (as an interval, to allow for potential measurement error). We add these numbers to arrive at a total annual figure for GSE subprime PLS purchases from 2003-2007. In the last column of the table we show the annual purchases derived from our algorithm, which closely tracks the purchases derived from the public sources.

Although a complete list of PLS securities purchased by Fannie Mae and Freddie Mac is not

[23]In our sample of pools backing securities issued between 2003 and 2007 there are 245 pools for which the share of second lien loans is greater than 75 percent.

[24]We also imposed the restriction that a GSE pool can only be associated with a deal that contained at least two mortgage pools. We did this because of our focus on deal-level fixed effects, however there is nothing that prohibited an issuer from structuring a deal with only a single conforming loan pool. There were only a handful of these deals in the CoreLogic database, so the restriction has no effect on the algorithm.

Table A.2: Loan-level Summary Statistics: Corelogic Subprime Low Documentation PLS Issued 2003–2007

	FHFA Total	Report to Congress Fannie Mae	FCIC Report Freddie Mac	FHFA + FCIC Reports Total	Algorithm Total
2003	.	25.8	[44-48]	[69.8-73.8]	67.7
2004	.	67.0	[70-74]	[137-141]	141.0
2005	.	24.4	[112-116]	[136.4-140.4]	134.4
2006	110.4	35.6	[72-76]	[107.6-111.6]	106.0
2007	59.6	16.0	[37-41]	[53-57]	50.1

publicly available, one source of information for validating our algorithm at the security-level is a disclosure by the Federal Housing Finance Agency (FHFA) announcing lawsuits against PLS issuers in September of 2011.[25]

In the lawsuits, the FHFA focuses on 718 securities that were purchased by the GSEs, and includes the associated tickers (e.g. "ABFC 2006-HE1 A1"). We match the 718 securities to subprime PLS tickers obtained from Bloomberg. We are able to identify 478 out of the 718 securities as being subprime, while another 226 have different collateral characteristics.[26] The face values of the 478 subprime securities included in the lawsuit were $37.3 billion in 2005, $80.7 billion in 2006, and $38.3 billion in 2007, vastly less than the total amount of subprime PLS purchased by the GSEs during those years.

We can use the 478 securities to partially validate our algorithm, as we are certain that these securities were purchased by the GSEs. When we match the securities to our GSE indicator variable we find that 476 out of the 478 subprime securities included in the lawsuit (99.6 percent) are classified as GSE, which translates into a type I error rate of 0.4 percent. We

[25]In this lawsuit, the FHFA sued 17 PLS issuers because it concluded that "some portion of the losses that Fannie Mae and Freddie Mac incurred on private-label mortgage-backed securities (PLS) are attributable to misrepresentations and other improper actions by the firms and individuals named in these filings." (FHFA, September 6, 2011, "Federal Housing Finance Agency Statement on Recent Lawsuits Filed")

[26]This leaves out 14 securities that we are not able to match to Bloomberg using the tickers provided in the lawsuit documents.

cannot use this test to evaluate the type II error in our algorithm, given that there are many securities purchased by Fannie Mae and Freddie Mac that are not part of the lawsuits.

A.3 Robustness Tables

This section contains tables of the robustness checks referred to in the main text.

Table A.3 displays summary statistics (from Bloomberg) for the triple-A securities that are collateralized by the loans in the CoreLogic database broken down by whether the securities are derived from GSE or non-GSE mortgage pools. For each year, the table shows the number of associated mortgage pools, the average size of the mortgage pools, the spread between the average coupon of the triple-A securities and the one-month LIBOR, and the weighted average expected life of the associated triple-A tranches using the sizes of each individual security as the weights. The last column in the table displays the differences in the summary statistics between the pool types and whether those differences are statistically different from zero. The average pool size is between \$19 and \$20 million for both GSE and non-GSE pools in each year of our sample. The average spread over one-month LIBOR on GSE triple-A securities was between two and six basis points higher than the average spread on the non-GSE triple-A securities in each year of our sample. Finally, the GSE and non-GSE pools have similar weighted average lives (around 2.4 to 2.5 years), where the life of each security is taken from the prospectus and is based on predicted prepayment behavior on the part of the borrowers.

Tables A.4 displays results broken down by issuer and originator affiliation for the matched sample of loans discussed in Section 4.4 in the text. The table is consistent with Table 5 in the text as it shows that the results are significantly stronger for the sample of deals in which the issuer and originator were affiliated.

Table A.5 displays results for the primary empirical specifications in the paper with the inclusion of zip code fixed effects (instead of state fixed effects). This means that all coefficients are estimated using variation within very small geographic areas, and any unobserved shocks that occur at this geographic level are absorbed by these fixed effects. Given how long it takes to estimate this model we use a 50 percent random sample of the loan-level data. The results are all economically and statistically very close to those in the previous tables, confirming that the GSE effect is not just picking up unobserved geographic differences in the distribution of loans along any of the dimensions mentioned above. The differential performance of GSE pools is concentrated in the sample of low-documentation loans, and the GSE Deal Fraction variable is associated with better performance of loans in GSE pools. Table A.6 shows the split between loans from affiliated and unaffiliated issuer-originator pairs, and, as before, all of the effect of GSE Deal Fraction is concentrated in deals where originators and issuers are affiliated.

Tables A.7, A.8, and A.9 correspond to a robustness check where we focus exclusively on the sample of low-documentation mortgages only. This allows us to address the concern that low-documentation loans might be differentially sensitive to observable risk characteristics, in particular those in GSE pools, and that, by pooling all loans in the sample, we might confound the GSE effect with this differential sensitivity. Table A.7 displays summary statistics for the sample of low-documentation mortgages. The table is structured exactly like Table 2 in the text and shows both unconditional and conditional (on deal fixed effects) summary statistics for the full sample of low-documentation loans in addition to summary statistics for the matched sample of low-documentation mortgages. The patterns are very similar to those found in Table 2 . There are relatively large differences in the unconditional covariate means between loans in GSE and non-GSE pools. These differences, in most cases, shrink when we condition on deal fixed effects, and are further reduced in the matched sample. Generally, while the means of each variable are different (for example, FICO scores are higher in the low-documentation sample, as are interest rates at origination), the differences between GSE and non-GSE pools are not systematically higher or lower compared to the full sample. Put differently, we do not find that the gap between the two types of pools is consistently larger or smaller for the low-documentation loans, and the differences are almost all in the same direction as in the full sample.

Table A.8 displays results when we estimate our primary specifications on the sample of low-documentation loans only with the same structure as Table 6 in the text and Table A.5 above, although, given that the sample includes only low-documentation loans, we cannot identify the low-documentation indicator variable separately. Also, columns (3), (6), and (9) include zip code fixed effects to control for local factors discussed above. All results are consistent for this subsample of loans, which again supports the hypothesis that private information on the part of securities issuers is driving the performance differences. We show results for the splits by affiliated and unaffiliated deals in Table A.9.

Table A.10 displays results when we exclude jumbo loans from the sample (i.e. loans with balances above the CLL). To the extent that jumbo loans might behave systematically worse due to unobserved characteristics, this could explain some of the results above.[27] The table shows that eliminating jumbo loans from the sample has little impact on the results.

Table A.11 displays results when we exclude second liens from the sample. GSE pools include significantly fewer second lien loans relative to non-GSE pools, so we are concerned that this could be driving some of the performance differences. The table shows that eliminating second liens has no impact on the results.

[27]This is really only a concern for the full, unmatched sample, since the vast majority of jumbo loans are dropped in the matching exercise. Jumbo loans comprise only 1.2 percent of loans in the matched sample.

Table A.12 displays results when we exclude fixed-rate mortgages from the sample and include only ARMs. This exercise is motivated by a concern that the results could be driven by potential differences in exposure to prepayment risk across GSE and non-GSE pools. In particular, it could be that the GSEs looked for a different pre-payment profile from their PLS securities than other investors, and that those differences in pre-payment profiles simply happened to be correlated with better credit risk ex-post. Since prepayment risk is really only an important issue with respect to FRMs, by restricting the sample to only ARMs we are able to keep prepayment profiles largely fixed, which indicates that results are not driven by unobserved credit factors that could be correlated with prepayment risk.

Tables A.13–A.17 contain estimation results for an additional set of robustness checks. Table A.13 displays results for an alternative default definition − 90+ days delinquent rather than 60+ days delinquent. The results are quite similar and suggest that they are not sensitive to the definition of default. Table A.14 displays results for two different horizons (24 months and 36 months) measured relative to the month of security issuance rather than a specific point in (calendar) time. These horizons are consistent with the methodology used in estimating the ex-ante default probabilities (Table 7 in the text). The estimates in Table A.14 are qualitatively and quantitatively similar to those in Tables 3 and 5 in the text.

In Table A.15 we re-estimate our ex-post default rate regressions using logistic models rather than linear probability models. The drawback of using logit models in the presence of fixed effects is the well-known incidental parameters problem. With small numbers of observations within groups, the incidental parameters problem can result in significant bias of the estimates of the slope parameters. Since most of the deals contain thousands of loans in multiple mortgage pools, this is likely not an important issue in our context. In fact, the results displayed in Table A.19 are virtually identical to the results in Tables 3 and 5 in the text.[28]

In Table A.16 we use an alternative method to calculate standard errors. Specifically, we cluster by deal rather than quarter of issuance. Since there are only 20 quarters in the sample, one might be concerned that clustering by quarter is producing a downward bias in the estimates of the standard errors. Clustering by deal is a natural alternative since the deal is one of the primary units of observation and all loans/securities in a deal share many common characteristics. Since there are almost 2,000 deals in the sample, clustering by deal addresses the potential downward bias due to a small number of clusters. As the table clearly shows, clustering by deal leads to significantly lower standard errors compared to clustering by quarter.

Table A.17 displays results in which we include a series of indicator variables that measure the fraction of the population in a zip code that resides in census tracts which meet the qualifications for the underserved area affordable housing goal (UAG). In order to determine underserved area goal (UAG) eligibility, we match tract-level UAG data obtained from the FHFA to the zip code associated with each mortgage using a population-weighted bridge provided by the Missouri Census Data Center. We use 1990 tract definitions for tract data up to 2002, and 2000 tract definitions for the later years. We include a zip-code level variable that varies between 0 if the zip code does not contain a census tract that is eligible for the UAG, and 1 if the entire population in that the zip code is in UAG-eligible census tracts. This variable takes intermediate values when only a fraction of the zip code's population is in such tracts. We view these results as supplementary to the specifications in Table A.5 (inclusion of zip code fixed effects), which provide a more powerful way of controlling for regional differences between loans, including UAG eligibility.

[28] We also considered the conditional logit estimator developed by Chamberlain (1980), which eliminates the fixed effects from the likelihood function, and thus is not susceptible to the incidental parameters problem. However, with large numbers of observations within groups (in our case loans within deals), the estimator becomes difficult to implement computationally.

Table A.18 displays summary statistics on the percentage of deals with missing information on loan servicer identity, and also the share of loans that are in deals with only one servicer. 80% of loans have servicer information and, of these, 80% are included in deals where all loans share the same servicer. In Table A.19 we show that the main results in the paper are not affected if we run regressions dropping all loans that are in deals with more than one servicer. In the remaining loans (about 6.2 million), the deal fixed effect account for fixed characteristics of the servicers.

Tables A.20–A.22 contain ex-ante predicted default probability results (corresponding to Table 7 in the text) in which predicted default probabilities are calculated using an alternative definition of default (90+ days delinquent) and different statistical models. Table A.20 shows that the alternative default definition does not affect the ex-ante results. Table A.21 displays results in which separate models for first and second lien mortgages and separate models for loans below and above the CLL are used, while Table A.22 shows results in which separate models for adjustable-rate and fixed-rate mortgages are utilized. The results are quite similar to those in Table 7, in which a single model was used to calculate all predicted default probabilities.

In the first column of Table A.23 we display output from our main specification (Table 3) for (virtually) all covariates, in order to provide the reader with an idea of the quantitative magnitudes of the estimates associated with the control variables in our regressions. In the second column, we display a similar set of estimates from the regression specification that includes controls for the fraction of the zip code (in which the loan was originated) that lies in census tracts that are eligible for the underserved area affordable housing goal (UAG).[29] We specify the variable has a set of indicators that correspond to each decile. For example,

[29] Both regressions also include a full set of state fixed-effects, deal fixed-effects, and dummy variables that control for missing covariate values.

the first variable is an indicator for whether the UAG zip code fraction is between 0 and 0.1, the second is an indicator for whether the UAG fraction is between 0.1 and 0.2, etc. We omit the indicator that corresponds to the highest UAG values (between 0.9 and 1). The estimation results show that loans originated in zip codes with higher UAG fractions are more likely to default, *ceteris paribus*.

Table A.24 displays results for the comparison of loans in non-GSE pools in deals with and without GSE participation. The table is structured exactly like Panel A of Table 9 in the main text, but shows estimation results for all three default horizons.

Figure A.1 displays the results from estimating our primary specifiation (Table 3) separately by quarter of issuance. Specifically, we plot the coefficient estimates associated with the GSE pool indicator variable and the its interaction with the low-documentation dummy variable for each quarter in our sample. The figure clearly shows that the results are not driven by one or two quarters alone, but rather that the effects of both the GSE dummy and the low-documentation interaction are always below zero, and become especially strong after the first quarter of 2005. The figure is created by estimating the regressions separately for each quarter, which is equivalent to including a full set of interactions between all of the controls and quarter of issuance indicator variables. This also eliminates concerns related to non-linearities that might be induced by observable differences over time between GSE and non-GSE pools, and that might persist even after accounting for the covariates in the main (pooled) models. In Section 4.5 we also show that results are unchanged when we restrict the analysis to low-documentation loans only (which allows the coefficients on all observable characteristics to be estimated on this sample of loans only).

Figure A.2 shows the estimated propensity score distributions for loans in GSE and non-GSE pools from the matching exercise described in Section 4.4 in the paper. We include the same set of covariates in the propensity score model as in the main regression models. The figure shows that there is substantial overlap in the propensity score distributions, particularly between the values of 0.3 and 0.8.

Table A.3: Pool-level Summary Statistics: Corelogic Subprime PLS Issued 2003–2007

Year		Non-GSE	GSE	Difference
2003	# Pools	312	172	140
	Pool Size ($ millions)	19.34	19.83	0.49***
	Spread (bps)	36.96	38.83	1.87
	Average Life (years)	2.91	2.85	-0.07
2004	# Pools	419	297	122
	Pool Size ($ millions)	19.59	20.11	0.52***
	Spread (bps)	30.16	33.17	3.01***
	Average Life (years)	2.66	2.76	0.10*
2005	# Pools	511	316	195
	Pool Size ($ millions)	19.92	20.02	0.11**
	Spread (bps)	20.02	25.88	5.86***
	Average Life (years)	2.31	2.51	0.19***
2006	# Pools	537	314	223
	Pool Size ($ millions)	20.05	19.72	-0.32***
	Spread (bps)	13.46	16.44	2.98***
	Average Life (years)	2.15	2.30	0.15***
2007	# Pools	241	171	70
	Pool Size ($ millions)	19.90	19.58	-0.32***
	Spread (bps)	23.47	25.27	1.80
	Average Life (years)	2.20	2.18	-0.02
All	# Pools	2,020	1,270	750
	Pool Size ($ millions)	19.79	19.88	0.09***
	Spread (bps)	23.41	26.92	3.51***
	Average Life (years)	2.42	2.51	0.09***

Notes: This table shows summary statistics for triple-A subprime PLS issued between 2003 and 2007 broken down by whether the security was collateralized by GSE or non-GSE mortgage pools. GSE refers to mortgage pools made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. The spread refers to the difference between the average coupon of triple-A subprime securities (weighted by the size of the tranche in each pool) and the one-month LIBOR. The average life refers to the average expected life for the tranches as advertised in the prospectus (where the average for the pools was weighted by the size of each tranche). Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.4: Default, "GSE Deal Fraction," and Issuer/Originator Affiliation: Matched Sample

	All Deals			Affiliated Deals			Unaffiliated Deals		
Horizon	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4
GSE (d)	0.008***	0.005*	0.005	0.019**	0.015**	0.013**	-0.008	-0.009	-0.009
	(2.35)	(1.66)	(1.60)	(2.15)	(2.10)	(2.06)	(1.18)	(1.35)	(1.51)
GSE*Low Doc	-0.019***	-0.014***	-0.012***	-0.025***	-0.015***	-0.012***	-0.018***	-0.013***	-0.013***
	(9.34)	(8.39)	(7.33)	(5.81)	(4.54)	(3.76)	(7.53)	(9.04)	(8.81)
GSE*"GSE Deal Fraction"	-0.031***	-0.026***	-0.024***	-0.045***	-0.039***	-0.037***	-0.009	-0.008	-0.007
	(5.39)	(4.34)	(4.34)	(3.00)	(3.03)	(3.18)	(0.92)	(0.95)	(0.93)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	3,383,356	3,383,356	3,383,356	956,586	956,586	956,586	1,180,333	1,180,333	1,180,333
# Deals	1,099	1,099	1,099	311	311	311	453	453	453
Adjusted R^2	0.14	0.19	0.20	0.14	0.18	0.20	0.15	0.19	0.20

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on a matched sample of loans in GSE and non-GSE pools. The details of the construction of the matched sample are discussed in section 4.2 of the text. The first three columns display results for the sample of all deals, while columns 4-6 display results where the originator of all loans in a deal is affiliated with the issuer, while the last three columns display results for the sample of deals in which the originator of all loans in a deal is not affiliated with the issuer. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.5: Effect of GSE Participation on Default: Zip Code Fixed Effects

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GSE (d)	-0.011***	-0.002	0.017***	-0.013***	-0.005*	0.007*	-0.014***	-0.007***	0.002
	(7.31)	(0.91)	(4.81)	(6.58)	(1.86)	(1.66)	(6.86)	(2.74)	(0.36)
GSE*Low Doc		-0.027***	-0.027***		-0.024***	-0.024***		-0.021***	-0.022***
		(8.13)	(8.04)		(8.05)	(8.37)		(8.16)	(8.61)
GSE*"GSE Deal Fraction"			-0.029***			-0.017***			-0.013*
			(5.21)			(2.45)			(1.82)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Zip Code F.E.?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	5,226,211	5,226,211	5,073,048	5,226,211	5,226,211	5,073,048	5,226,211	5,226,211	5,073,048
# Deals	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809
Adjusted R^2	0.19	0.19	0.19	0.23	0.23	0.23	0.24	0.24	0.24

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions include a full set of zip code fixed effects. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). This regression is estimated on a 50 percent random sample of the data. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.6: Default, "GSE Deal Fraction," and Issuer/Originator Affiliation: Zip Code Fixed Effects

	All Deals			Affiliated Deals			Unaffiliated Deals		
Horizon	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4
GSE (d)	0.017***	0.007*	0.002	0.017*	0.016**	0.014**	-0.005	-0.009	-0.011**
	(4.81)	(1.66)	(0.36)	(1.78)	(2.16)	(2.01)	(0.98)	(1.59)	(2.15)
GSE*Low Doc	-0.027***	-0.024***	-0.022***	-0.028***	-0.019***	-0.015***	-0.024***	-0.018***	-0.016***
	(8.04)	(8.37)	(8.61)	(5.30)	(4.42)	(3.89)	(5.92)	(4.50)	(4.18)
GSE*"GSE Deal Fraction"	-0.029***	-0.017***	-0.013*	-0.035***	-0.033***	-0.030***	-0.001	0.004	0.007
	(5.21)	(2.45)	(1.82)	(2.58)	(3.12)	(3.02)	(0.15)	(0.48)	(0.88)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	5,073,048	5,073,048	5,073,048	1,329,705	1,329,705	1,329,705	1,687,168	1,687,168	1,687,168
# Deals	1,809	1,809	1,809	396	396	396	695	695	695
Adjusted R^2	0.19	0.23	0.24	0.18	0.22	0.23	0.18	0.23	0.24

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on a matched sample of loans in GSE and non-GSE pools. The details of the construction of the matched sample are discussed in section 4.2 of the text. The first three columns display results for the sample of all deals, while columns 4-6 display results where the originator of all loans in a deal is affiliated with the issuer, while the last three columns display results for the sample of deals in which the originator of all loans in a deal is not affiliated with the issuer. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.7: Loan-level Summary Statistics: Corelogic Subprime Low Documentation PLS Issued 2003–2007

	Full sample (unconditional)		Full sample (conditional on deal F.E.)		Matched Sample	
	Non-GSE N = 2,591,963	GSE N = 1,430,083	Non-GSE N = 2,591,963	GSE N = 1,430,083	Non-GSE N = 627,358	GSE N = 589,164
Continuous Variables	Mean	Mean	Mean	Mean	Mean	Mean
FICO (Points)	662	632	655	644	640	642
Balance ($)	171,898	172,915	199,280	123,266	186,999	173,902
CLTV (P.Points)	89.3	83.6	88.5	85.1	87.4	85.4
Orig. Rate (P. Points)	9.01	8.11	8.71	8.65	8.41	8.30
Term (months)	307	350	320	325	346	343
Unemployment (P. Points)	4.97	5.36	5.08	5.17	5.15	5.10
Trailing 12-month unemployment change	-7.6%	-6.1%	-7.3%	-6.7%	-7.4%	-7.0%
Unemployment change through 2012	72.6%	61.0%	69.8%	66.0%	69.7%	66.8%
Price Index	195	177	193	182	191	185
Trailing 12-month HPA	13.0%	13.3%	13.4%	12.5%	13.2%	13.2%
HPA through 2012	-21.4%	-15.7%	-19.6%	-18.9%	-20.4%	-18.7%
UAG %	51.5%	54.8%	51.9%	55.8%	56.5%	55.7%
Indicator Variables						
Low Documentation (share)	1	1	1	1	1	1
Non-Owner Occupied (share)	0.113	0.120	0.100	0.143	0.087	0.116
Purchase Loan (share)	0.602	0.401	0.597	0.411	0.611	0.489
Cash-Out Refinance (share)	0.345	0.537	0.353	0.522	0.349	0.453
Interest-Only (share)	0.160	0.087	0.155	0.092	0.126	0.119
Balloon (share)	0.279	0.118	0.228	0.212	0.180	0.162
ARM (share)	0.480	0.772	0.553	0.640	0.716	0.689
Prepay Penalty (share)	0.576	0.706	0.610	0.602	0.709	0.679
Default Rate through 2008Q4	0.383	0.331	0.382	0.334	0.435	0.367
Default Rate through 2010Q4	0.473	0.393	0.459	0.418	0.508	0.447
Default Rate through 2012Q4	0.495	0.409	0.478	0.440	0.524	0.467

Notes: This table shows summary statistics for both the full sample of *low-documentation* loans underlying the GSE and non-GSE triple-A subprime PLS securities in the Corelogic database issued between 2003 and 2007 as well as in a matched sample of *low-documentation* loans in GSE and non-GSE loan pools. The details of the construction of the matched sample are discussed in section 4.2 of the text. The table displays both unconditional full sample means and full sample means conditional on deal fixed effects. Since the matched sample is constructed within deals, the matched sample summary statistics are also conditional on deal fixed effects.

Table A.8: Effect of GSE Participation on Default: Low Documentation Sample

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
GSE (d)	-0.025***	0.005	0.010***	-0.020***	0.003	-0.001	-0.020***	0.003	-0.006*
	(11.10)	(1.35)	(2.63)	(9.26)	(1.12)	(0.38)	(8.65)	(1.04)	(1.95)
GSE*"GSE Deal Fraction"		-0.046***	-0.044***		-0.036***	-0.026***		-0.035***	-0.021***
		(7.08)	(7.63)		(6.06)	(4.88)		(5.80)	(5.03)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Zip Code F.E.?	N	N	Y	N	N	Y	N	N	Y
# Loans	4,021,713	3,893,223	3,889,065	4,021,713	3,893,223	3,889,065	4,021,713	3,893,223	3,889,065
# Deals	1,782	1,707	1,707	1,782	1,707	1,707	1,782	1,707	1,707
Adjusted R^2	0.22	0.22	0.22	0.26	0.26	0.26	0.27	0.27	0.27

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on only the sample of low-documentation mortgages in GSE and non-GSE pools. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 3. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.9: Default, “GSE Deal Fraction,” and Issuer/Originator Affiliation: Low Documentation Sample

	All Deals			Affiliated Deals			Unaffiliated Deals		
Horizon	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4	2008:Q4	2010:Q4	2012:Q4
GSE (d)	0.005	0.003	0.003	0.020	0.016	0.017^{*}	-0.017^{*}	-0.017^{*}	-0.018^{**}
	(1.35)	(1.12)	(1.04)	(1.33)	(1.46)	(1.68)	(1.88)	(1.90)	(2.08)
GSE*“GSE Deal Fraction”	-0.046^{***}	-0.036^{***}	-0.035^{***}	-0.072^{***}	-0.061^{***}	-0.061^{***}	-0.015	-0.012	-0.010
	(7.08)	(6.06)	(5.80)	(3.04)	(3.23)	(3.47)	(1.23)	(0.95)	(0.80)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	3,893,223	3,893,223	3,893,223	858,256	858,256	858,256	1,405,490	1,405,490	1,405,490
# Deals	1,707	1,707	1,707	390	390	390	691	691	691
Adjusted R^2	0.19	0.23	0.24	0.18	0.23	0.24	0.18	0.24	0.25

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on only the sample of low/no documentation loans in GSE and non-GSE pools. The first three columns display results for the sample of all deals, while columns 4-6 display results where the originator of all loans in a deal is affiliated with the issuer, while the last three columns display results for the sample of deals in which the originator of all loans in a deal is not affiliated with the issuer. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is “GSE” which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. “GSE Deal Fraction” is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with covariates include controls for a large number of borrower and loan characteristics (a full list is given in Section 2.1). Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.10: Default for Loans in GSE and Non-GSE Pools: Loans Below the CLL Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.022***	-0.010***	0.011***	-0.019***	-0.007***	0.011***	-0.018***	-0.007***	0.011***
	(10.67)	(5.69)	(2.87)	(8.81)	(3.07)	(3.48)	(8.04)	(2.94)	(3.66)
GSE*Low Doc		-0.034***	-0.033***		-0.032***	-0.033***		-0.030***	-0.030***
		(9.28)	(9.07)		(9.40)	(9.40)		(9.58)	(9.56)
GSE*"GSE Deal Fraction"			-0.034***			-0.029***			-0.028***
			(5.85)			(4.64)			(4.44)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	9,783,310	9,783,310	9,495,412	9,783,310	9,783,310	9,495,412	9,783,310	9,783,310	9,495,412
# Deals	1,809	1,809	1,724	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.15	0.15	0.15	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Loans above the conforming loan limit (jumbo mortgages) are excluded from the sample.

Table A.11: Default for Loans in GSE and Non-GSE Pools: First Lien Mortgages Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.014***	-0.006***	0.013***	-0.013***	-0.006***	0.011***	-0.013***	-0.006***	0.010***
	(7.28)	(2.91)	(3.46)	(6.37)	(3.11)	(3.80)	(5.96)	(3.15)	(3.88)
GSE*Low Doc		-0.026***	-0.026***		-0.020***	-0.020***		-0.018***	-0.018***
		(9.48)	(8.99)		(8.11)	(7.89)		(7.47)	(7.30)
GSE*"GSE Deal Fraction"			-0.030***			-0.027***			-0.026***
			(4.83)			(4.53)			(4.50)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue Year F.E. ?	.	.	.	.	.	.	.	.	.
# Loans	7,743,382	7,743,382	7,530,399	7,743,382	7,743,382	7,530,399	7,743,382	7,743,382	7,530,399
# Deals	1,632	1,632	1,561	1,632	1,632	1,561	1,632	1,632	1,561
Adjusted R^2	0.14	0.14	0.14	0.19	0.19	0.18	0.19	0.19	0.19

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Second lien mortgages are excluded from the sample.

Table A.12: Default for Loans in GSE and Non-GSE Pools: Adjustable-Rate Mortgages Only

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.015***	-0.005**	0.019***	-0.015***	-0.007***	0.015***	-0.015***	-0.007***	0.013***
	(7.23)	(2.04)	(5.19)	(6.76)	(2.62)	(4.72)	(6.61)	(2.90)	(4.39)
GSE*Low Doc		-0.028***	-0.028***		-0.023***	-0.023***		-0.021***	-0.021***
		(8.92)	(8.54)		(7.87)	(7.69)		(7.51)	(7.32)
GSE*"GSE Deal Fraction"			-0.037***			-0.033***			-0.032***
			(6.06)			(5.59)			(5.43)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	6,161,367	6,161,367	5,971,766	6,161,367	6,161,367	5,971,766	6,161,367	6,161,367	5,971,766
# Deals	1,634	1,634	1,557	1,634	1,634	1,557	1,634	1,634	1,557
Adjusted R^2	0.14	0.15	0.14	0.2	0.2	0.19	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Fixed-rate mortgages are excluded from the sample.

Table A.13: Default for Loans in GSE and Non-GSE Pools: Alternative Default Definition

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.019***	-0.007***	0.012***	-0.016***	-0.005***	0.009***	-0.015***	-0.006***	0.008***
	(10.43)	(4.66)	(3.45)	(9.62)	(2.68)	(3.18)	(8.75)	(2.65)	(3.21)
GSE*Low Doc		-0.033***	-0.033***		-0.029***	-0.030***		-0.027***	-0.027***
		(9.88)	(9.66)		(8.91)	(9.07)		(8.76)	(8.95)
GSE*"GSE Deal Fraction"			-0.031***			-0.023***			-0.021***
			(5.41)			(3.84)			(3.66)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202
# Deals	1,809	1,809	1,724	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.15	0.15	0.15	0.21	0.21	0.21	0.22	0.22	0.22

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 90 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.14: Default for Loans in GSE and Non-GSE Pools: Alternative Horizons

Horizon		24 Months			36 Months	
GSE (d)	-0.016***	-0.006***	0.008***	-0.016***	-0.005***	0.012***
	(6.89)	(3.26)	(1.77)	(8.50)	(2.42)	(3.27)
GSE*Low Doc		-0.027***	-0.028***		-0.031***	-0.031***
		(6.78)	(7.08)		(9.39)	(9.41)
GSE*"GSE Deal Fraction"			-0.021***			-0.027***
			(2.83)			(4.46)
Deal F.E. ?	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y
# Loans	10,464,165	10,464,165	10,156,202	10,464,165	10,464,165	10,156,202
# Deals	1,809	1,809	1,724	1,809	1,809	1,724
Adjusted R^2	0.16	0.16	0.16	0.20	0.20	0.20

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated over two different horizons relative to the month of security issuance: 24 months and 36 months. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. This regression was estimated on a 50 percent random sample of the data. Standard errors are heteroskedasticity-robust and clustered by quarter of issue. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.15: Default for Loans in GSE and Non-GSE Pools: Logit Model

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.015***	-0.004***	0.016***	-0.014***	-0.005***	0.013***	-0.014***	-0.005***	0.012***
	(10.70)	(2.56)	(4.56)	(9.09)	(2.53)	(4.15)	(8.91)	(2.87)	(3.81)
GSE*Low Doc		-0.030***	-0.030***		-0.027***	-0.027***		-0.025***	-0.025***
		(9.08)	(8.95)		(7.75)	(7.84)		(7.48)	(7.60)
GSE*"GSE Deal Fraction"			-0.032***			-0.027***			-0.026***
			(5.82)			(4.55)			(4.32)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	10,018,355	10,018,355	9,742,002	10,018,355	10,018,355	9,742,002	10,018,355	10,018,355	9,742,002

Notes: This table shows average partial effects from loan-level, logistic regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 60 days or more delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the average partial effects, the second row shows z-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.16: Default for Loans in GSE and Non-GSE Pools: Clustering at Deal-level

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.019***	-0.007***	0.019***	-0.016***	-0.006***	0.007**	-0.015***	-0.006***	0.002
	(16.01)	(5.33)	(5.25)	(14.44)	(4.11)	(2.20)	(14.03)	(4.33)	(0.57)
GSE*Low Doc		-0.032***	-0.032***		-0.029***	-0.029***		-0.026***	-0.027***
		(17.64)	(17.66)		(13.88)	(14.39)		(12.93)	(13.41)
GSE*"GSE Deal Fraction"			-0.037***			-0.023***			-0.018***
			(6.53)			(4.48)			(3.46)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issue Year F.E. ?	.	.	.	.	.	.	.	.	.
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	10,464,165	10,464,165	10,464,165	10,464,165	10,464,165
# Deals	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809	1,809
Adjusted R^2	0.16	0.16	0.16	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 20010Q4 and 2012Q4. Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 4. Standard errors are heteroskedasticity-robust and clustered at the deal level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.17: Robustness Check: Controlling for Affordable Housing Goal Effects

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.019***	-0.008***	0.013***	-0.016***	-0.006***	0.011***	-0.015***	-0.006***	0.011***
	(11.13)	(4.73)	(3.29)	(9.54)	(2.87)	(3.67)	(8.61)	(2.83)	(3.67)
GSE*Low Doc		-0.032***	-0.032***			-0.029***		-0.026***	-0.026***
		(9.04)	(8.90)			(8.53)		(8.43)	(8.44)
GSE*"GSE Deal Fraction"			-0.032***			-0.027***			-0.026***
			(5.53)			(4.46)			(4.21)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	9,823,431	9,823,431	9,554,443	9,823,431	9,823,431	9,554,443	9,823,431	9,823,431	9,554,443
# Deals	1,809	1,809	1,691	1,809	1,809	1,691	1,809	1,809	1,691
Adjusted R^2	0.16	0.16	0.16	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008Q4, 2010Q4 and 2012Q4. In addition to the full list of those controls discussed in section 4 of the text, we include a series of indicator variables that measure the fraction of the population in a zip code that resides in census tracts which meet the qualifications for the underserved area affordable housing goal (UAG). Default is defined as a loan being 60 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.18: Summary Statistics on Servicer Heterogeneity

	# Deals (2003-2007)	% of Deals
Servicer ID is Populated	1,455	80.43
Servicer ID is Missing for All Loans in Deal	318	17.58
Servicer ID is Missing for Some Loans in Deal	36	1.99
Total	1,809	100

	# Deals (2003-2007)	% of Deals
Same Servicer for All Loans in Deal	1,170	80.41
Different Servicers in Deal	285	19.59
Total	1,455	100

Table A.19: Default for Loans in GSE and Non-GSE Pools: Deals with Same Servicer

	Horizon through 2008:Q4			Horizon through 2010:Q4			Horizon through 2012:Q4		
GSE (d)	-0.021***	-0.009***	0.007*	-0.017***	-0.008***	0.006	-0.017***	-0.009***	0.005
	(10.74)	(4.50)	(1.70)	(7.23)	(3.17)	(1.62)	(6.63)	(3.23)	(1.43)
GSE*Low Doc		-0.032***	-0.032***		-0.026***	-0.026***		-0.023***	-0.024***
		(7.52)	(7.23)		(6.77)	(6.79)		(6.56)	(6.59)
GSE*"GSE Deal Fraction"			-0.023***			-0.020***			-0.019***
			(3.48)			(3.28)			(3.29)
Deal F.E. ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	6,413,780	6,413,780	6,161,913	6,413,780	6,413,780	6,161,913	6,413,780	6,413,780	6,161,913
# Deals	1,169	1,169	1,106	1,169	1,169	1,106	1,169	1,169	1,106
Adjusted R^2	0.15	0.15	0.15	0.20	0.20	0.19	0.21	0.21	0.21

Notes: This table shows loan-level, OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. The regressions are estimated on only the sample of PLS deals in which all loans were handled by the same servicer. Default is defined as a loan being 60 days delinquent or more at any point since origination through each of the horizons shown. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. Regressions with "covariates" include controls for a large number of borrower and loan characteristics. A full list of those controls is given in the text in Section 3. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Table A.20: Predicted Default Probabilities for Loans in GSE and Non-GSE Pools: Alternative Default Definition

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0009	-0.0075***	0.0085***	-0.0067***	0.0103***	-0.0079***
	(0.48)	(5.96)	(2.47)	(3.32)	(2.71)	(3.77)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,808	1,808	1,808	1,808	1,570	1,570
R^2	0.00	0.40	0.00	0.33	0.00	0.30

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0033	-0.0050***	0.0108***	0.0024	0.0188***	0.0039***
	(1.39)	(2.88)	(3.35)	(1.55)	(7.14)	(2.36)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,808	1,808	1,806	1,806	1,555	1,555
Pseudo R^2	0.00	0.34	0.00	0.24	0.01	0.17

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	-0.0025	-0.0046***	0.0106***	0.0035**	0.0186***	0.0065***
	(1.11)	(2.72)	(3.14)	(2.21)	(7.85)	(3.99)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,808	1,808	1,808	1,808	1,570	1,570
Pseudo R^2	0.00	0.34	0.00	0.24	0.01	0.17

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month ex-ante default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month ex-ante rates. Default is defined as a loan being 90 days delinquent or more at any point since origination. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.21: Ex-Ante Default Probabilities for Loans in GSE and non-GSE Pools: Alternative Model for Generating Predicted Default Rates

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0006	-0.0034	0.0241***	0.0096***	0.0232***	0.0104***
	(0.11)	(1.07)	(6.25)	(3.31)	(8.72)	(2.55)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.35	0.01	0.22	0.01	0.19

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0044	-0.0023	0.0212***	0.0079***	0.0213***	0.0090***
	(0.92)	(0.82)	(6.08)	(4.38)	(5.71)	(3.90)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.32	0.01	0.18	0.01	0.19

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0035	-0.0026	0.0211***	0.0092***	0.0189***	0.0093***
	(0.74)	(0.91)	(6.04)	(5.59)	(5.39)	(3.53)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.33	0.01	0.18	0.01	0.16

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Predicted default rates are calculated by estimating separate regressions for first and second mortgages and separate regressions for conforming and non-conforming (jumbo) loans. Default is defined as a loan being 60 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.22: Ex-Ante Default Probabilities for Loans in GSE and Non-GSE Pools: Alternative Model for Generating Predicted Default Rates

Panel A: OLS (Linear Probability Model)

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0017	-0.0038*	0.0159***	-0.0052***	0.0045	-0.0092***
	(0.55)	(1.75)	(3.82)	(2.53)	(0.84)	(3.85)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,465,022	10,465,022	10,465,022	10,465,022	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
R^2	0.00	0.32	0.00	0.23	0.00	0.26

Panel B: Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0038	-0.0018	0.0214***	0.0064***	0.0197***	0.0057***
	(0.89)	(0.84)	(4.91)	(2.77)	(4.86)	(2.42)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,438,305	10,438,305	10,330,159	10,330,159	8,779,903	8,779,903
# Deals	1,809	1,809	1,807	1,807	1,556	1,556
Pseudo R^2	0.00	0.33	0.01	0.17	0.01	0.17

Panel C: Multinomial Logit

	12-month Horizon		24-month Horizon		36-month Horizon	
GSE (d)	0.0038	-0.0023	0.0219***	0.0054**	0.0182***	0.0046***
	(0.90)	(1.04)	(6.56)	(2.28)	(4.79)	(1.84)
Deal F.E.?	N	Y	N	Y	N	Y
# Loans	10,464,165	10,464,165	10,464,165	10,464,165	9,168,963	9,168,963
# Deals	1,809	1,809	1,809	1,809	1,571	1,571
Pseudo R^2	0.00	0.34	0.01	0.14	0.00	0.17

Notes: This table shows loan-level, OLS regressions where the dependent variables are the 12-month, 24-month, and 36-month predicted default rates at the time the loan is originated using all information in the data for the previous two years for the 12-month rate and three years for the 24-month and 36-month predicted rates. Predicted default rates are calculated by estimating separate regressions for adjustable-rate and fixed-rate mortgages. Default is defined as a loan being 60 days delinquent, in foreclosure or REO. The independent variable of interest is "GSE" which is a 0/1 indicator variable. GSE pools have claims on groups of mortgages made up of almost exclusively loans below the conforming loan limit, whereas non-GSE pools refer to mortgage pools made up of loans both above and below the conforming loan limit. Panel A computes predicted default rates using OLS regressions, while Panel B computes predicted default rates using logistic regressions, and Panel C computes predicted default rates using multinomial logistic regressions. Standard errors are heteroskedasticity-robust and clustered at the quarter of issuance level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A.23: Ex-post Default Rate Linear Probability Model Coefficient Estimates

	Horizon through 2008	
	Baseline Specification	Include UAG Controls
GSE (d)	-0.007	-0.008
	(4.17)	(4.73)
Low Doc (d)	0.070	0.068
	(10.04)	(10.16)
GSE*Low Doc	-0.032	-0.032
	(9.19)	(9.04)
Owner Occupant (d)	-0.047	-0.044
	(11.72)	(11.29)
Prepay Penalty (d)	0.047	0.046
	(9.47)	(9.49)
1-unit Single Family Prop. (d)	-0.004	-0.002
	(3.03)	(1.83)
Condominium (d)	-0.024	-0.019
	(10.71)	(9.00)
Balloon (d)	0.049	0.048
	(12.12)	(11.88)
# Months Seasoned	0.000	0.000
	(0.00)	(0.04)
ARM (d)	-0.003	0.031
	(0.08)	(0.88)
Interest-Only (d)	0.046	0.046
	(10.31)	(10.46)

Negatively Amortizing (d)	0.046	0.045
	(1.73)	(1.70)
First Lien (d)	0.037	0.024
	(2.54)	(1.56)
Purchase Loan (d)	0.012	0.012
	(3.23)	(3.41)
Refinance Cash-Out (d)	-0.017	-0.018
	(13.21)	(13.65)
LTV	0.002	0.002
	(8.53)	(8.28)
$70 \leq$ LTV < 80 (d)	0.023	0.024
	(5.04)	(5.11)
$80 <$ LTV < 90 (d)	0.047	0.048
	(5.20)	(5.30)
$90 \leq$ LTV < 100 (d)	0.074	0.075
	(6.37)	(6.49)
LTV ≥ 100 (d)	0.130	0.131
	(8.24)	(8.32)
LTV $= 80$ (d)	0.026	0.025
	(6.78)	(6.62)
FICO	-0.001	-0.001
	(25.93)	(26.03)
FICO < 580	0.025	0.026
	(7.05)	(7.39)
$580 <$ FICO ≤ 620	0.022	0.023
	(5.90)	(6.12)
$620 <$ FICO ≤ 660	0.004	0.004

	(1.24)	(1.54)
660 < FICO $\leq$ 700	-0.010	-0.009
	(4.32)	(4.12)
Interest Rate	0.030	0.029
	(17.04)	(17.29)
Log (Loan Balance)	0.020	0.027
	(2.27)	(2.90)
Term	0.000	0.000
	(8.73)	(8.55)
Jumbo (d)	0.023	0.025
	(4.85)	(5.37)
Unemp. Level at Origination	0.004	0.003
	(6.17)	(4.97)
Price Index Level at Origination	0.001	0.001
	(7.90)	(7.41)
Δ Unemp. through 2008	0.013	0.019
	(2.76)	(3.67)
HPA through 2008	-0.190	-0.191
	(4.30)	(4.38)
0 $\leq$ UAG Fraction < 0.10	.	-0.040
	.	(11.57)
0.10 $\leq$ UAG Fraction < 0.20	.	-0.038
	.	(11.31)
0.20 $\leq$ UAG Fraction < 0.30	.	-0.030
	.	(8.62)
0.30 $\leq$ UAG Fraction < 0.40	.	-0.029
	.	(10.60)

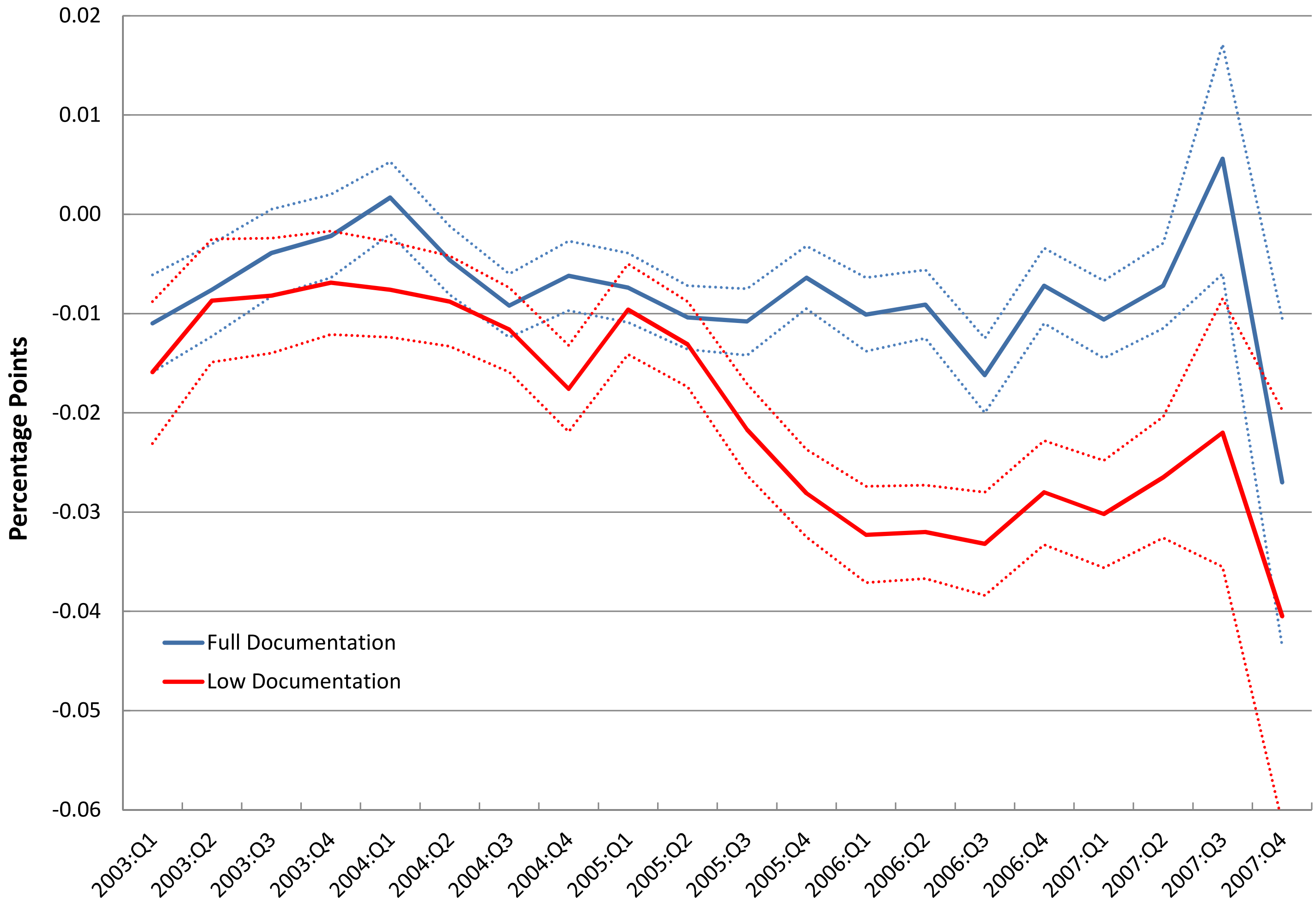
$0.40 \leq$ UAG Fraction < 0.50	.	-0.023
	.	(7.59)
$0.50 \leq$ UAG Fraction < 0.60	.	-0.026
	.	(9.04)
$0.60 \leq$ UAG Fraction < 0.70	.	-0.017
	.	(11.77)
$0.70 \leq$ UAG Fraction < 0.80	.	-0.014
	.	(7.04)
$0.80 \leq$ UAG Fraction < 0.90	.	-0.008
	.	(7.06)
Deal F.E. ?	Y	Y
State F.E. ?	Y	Y
# Loans	10,464,165	9,823,431
# Deals	1,809	1,809
Adjusted R^2	0.16	0.16

Table A.24: Comparison of Non-GSE Loans in GSE and Non-GSE Pools

Horizon (through)		2008			2010			2012	
GSE Deal	0.029***	0.024***	0.007	0.027***	0.029***	0.006	0.026***	0.029***	0.006
	(6.14)	(5.54)	(0.58)	(5.86)	(5.70)	(0.48)	(5.77)	(5.74)	(0.50)
GSE Deal * Low Doc		0.013***	0.012***		-0.004	-0.005		-0.008	-0.009**
		(2.92)	(2.86)		(0.99)	(1.31)		(1.87)	(2.24)
GSE Deal * Deal Fraction			0.030			0.041*			0.041**
			(1.63)			(2.28)			(2.40)
Covariates ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Issuer FEs ?	Y	Y	Y	Y	Y	Y	Y	Y	Y
# Loans	6,209,878	6,209,878	6,132,891	6,209,878	6,209,878	6,132,891	6,209,878	6,209,878	6,132,891
Adjusted R^2	0.16	0.16	0.17	0.20	0.20	0.20	0.21	0.21	0.21

Notes: This table displays results from a sample of mortgages in non-GSE, PLS pools only. The independent variable of interest is "GSE Deal" which is a 0/1 indicator variable that identifies loans in non-GSE mortgage pools that are in PLS deals with GSE participation. Panel A displays OLS regressions where the dependent variable is the default rate of loans backing subprime PLS issued between 2003 and 2007 calculated at three different points in time: 2008:Q4, 2010:Q4 and 2012:Q4. "GSE Deal Fraction" is a variable that is calculated quarterly for each deal sponsor that we are able to identify in the CoreLogic sample. It is calculated by taking the number of deals that include a GSE pool that the sponsor has issued up to each point in time and dividing that number by the total number of deals issued by the sponsor up until that quarter. All regressions include controls for a large number of borrower and loan characteristics including state and issue-year fixed effects (a full list is given in Section 2.1) as well as issuer fixed effects. Standard errors are heteroskedasticity-robust and clustered at the quarter of origination level. The first row for each variable shows the regression coefficient, the second row shows t-statistics. Level of statistical significance: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$

Figure A.1: Evolution of Ex-Post Performance Difference between GSE and Non-GSE Mortgage Pools over Sample Period



Notes: This figure plots the coefficient estimates of the GSE dummy variable and the interaction of the GSE dummy and the low-documentation dummy from a series of linear probability models estimated separately for each quarter of issuance over the sample period 2003–2007. The dependent variable is the default rate measured through 2008, where default is defined as a loan being 60 days delinquent or more at any point since origination. The model specification is identical to column 4 in Table 5, and is estimated separately for each quarter (of issuance) in the sample. The solid lines correspond to the point estimates from each linear probability model, while the 95 percent confidence intervals are given by the dashed lines.

Figure A.2: Estimated Propensity Score Distributions